PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12Q 1/68, 1/70, C12P 19/34, C12M 1/34, G01N 27/00, 21/64, C07H 21/02, 21/04, G06F 17/00

A1

(11) International Publication Number: WO 97/15690

(43) International Publication Date: 1 May 1997 (01.05.97)

(21) International Application Number: PCT/US96/17159

(22) International Filing Date: 24 October 1996 (24.10.96)

(30) Priority Data:

547,214	24 October 1995 (24.10.95)	US
663,823	14 June 1996 (14.06.96)	US

(81) Designated States: AL, AM, AU, AZ, BA, BB, BG, BR, BY, CA, CN, CU, CZ, EE, FI, GE, HU, IL, IS, JP, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, TJ, TM, TR, TT, UA, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(71) Applicant: CURAGEN CORPORATION [US/US]; 322 East Main Street, Branford, CT 06405 (US).

(72) Inventors: ROTHBERG, Jonathan, M.; 45B Cocheco Avenue, Branford, CT 06405 (US). DEEM, Michael, W.; 11136 Ophir Drive, Los Angeles, CA 90024 (US). SIMPSON, John, W.; Apartment B4, 23 Woodland Road, Madison, CT 06443 (US).

(74) Agents: ANTLER, Adriane, M. et al.; Pennie & Edmonds, 1155 Avenue of the Americas, New York, NY 10036 (US).

(54) Title: METHOD AND APPARATUS FOR IDENTIFYING, CLASSIFYING, OR QUANTIFYING DNA SEQUENCES IN A SAMPLE WITHOUT SEQUENCING

(57) Abstract

This invention provides methods by which biologically derived DNA sequences in a mixed sample or in an arrayed single sequence clone can be determined and classified without sequencing. The methods make use of information on the presence of carefully chosen target subsequences, typically of length from 4 to 8 base pairs, and preferably the length between target subsequences in a sample DNA sequence together with DNA sequence databases containing lists of sequences likely to be present in the sample to determine a sample sequence. The preferred method uses restriction endonucleases to recognize target subsequences and cut the sample sequence. Then carefully chosen recognition moieties are ligated to the cut fragments, the fragments amplified, and the experimental observation made. Polymerase chain reaction (PCR) is the preferred method of amplification. Several alternative embodiments are described which are capable of increased discrimination and which use TypeIIS restriction endonucleases, various capture moieties, or samples of specially synthesized cDNA. Another embodiment of the invention uses information on the presence or absence of carefully chosen target subsequences in a single sequence clone together with DNA sequence databases to determine the clone sequence. Computer implemented methods are provided to analyze the experimental results and to determine the sample sequences in question and to carefully choose target subsequences in order that experiments yield a maximum amount of information.

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]
1. Identify One or More Nucleic Acids in Sample Containing Two or More Nucleic Acids with which Nucleotide Sequences Differ. It is a classification or the approach of carrying out a quantum, and is the following process. : (a) The process which carries out the probe of the sample with one or more recognition means, However, each recognition means recognizes a different target nucleotide sub array or a different target nucleotide sub array of a set. (b) The process which makes one or more signals generate from this sample by which the probe was carried out with this recognition means, Each signal is generated from the nucleic acid in this sample. However, and the die length during existence of the target sub array in (i) this nucleic acid, And include the display of the identity of the target sub array in (ii) this nucleic acid, or the identity of this target sub array set (the target sub array in this nucleic acid is included in this). It reaches. (c) The process which searches a nucleotide sequence database in order to determine the nonexistence of the array which matches one or more generated signals, or all the arrays that match, However, this database includes many known nucleotide sequences of the nucleic acid which may exist in this sample. It has the die length during existence of the target sub array as what is expressed by the signal which carried out (i) generation with the same array from this database. And the same target sub array as what is expressed by the signal which carried out (ii) generation, When it has the target sub array which is expressed by the generated signal and which is the member of the same target sub array set, this array from this database matches the generated signal. Or by that cause one or more nucleic acids in this sample -- identification and a classification -- or a quantum is carried out The above-mentioned approach of coming to contain each process.
2. It is the approach according to claim 1 of matching the signal which generated the array from this database when it had the same target sub array as what is expressed by the signal generated while having the die length during existence of the same target sub array as what is expressed by the signal which each recognition means has recognized one target sub array, and the array from this database generated.
3. Approach according to claim 1 of being member of target sub array set to whom it is expressed by signal which recognizes target sub array each recognition means of whose is one set, and has die length during existence of the same target sub array as what is expressed by signal which array from this database generated, and this target sub array generated.
4. Approach according to claim 1 using one or more recognition means which take lessons from each part coming [dividing a nucleic-acid sample into two or more parts, and performing a process according to claim 1 separately to these two or more parts] further, and differ.
5. Approach according to claim 1 determined from quantitative level of one or more signals with which it was determined that abundance of nucleic acid including this nucleotide sequence in sample matched this array.
6. Approach according to claim 1 two or more nucleic acids are DNA.
7. Approach according to claim 6 DNA is cDNA.
8. Approach according to claim 7 by which cDNA is prepared from vegetation, unicellular animal, multicellular animal, bacteria, virus, fungus, or yeast.

9. Approach according to claim 8 said database includes all known manifestation arrays substantially [said vegetation, an unicellular animal, a multicellular animal, bacteria, a virus a fungus, or yeast].
10. The approach according to claim 7 cDNA is a thing from all the cells RNA or all cell Pori (A) RNA.
11. Recognition Means are One or More Restriction Endonucleases, and the Recognition Site is Target Sub Array. Said process which carries out a probe digests a sample by one or more restriction endonucleases. A fragment and nothing, It comes to contain connecting a double strand adapter DNA molecule with this fragment, and obtaining a connection fragment. However, each adapter DNA molecule consists of a part for the 1st and part II without (i) five prime end phosphoric-acid radical. A shorter chain (the amount of part I is complementary to the lobe which is in the five prime end of this short ****, and was generated by one of these the restriction endonucleases), (ii) Have a complementary three-dash terminal sub array in a part for part II of this short ****. a longer chain -- containing -- and -- Said generation process dissolves this short **** from a connection fragment. Contact this connection fragment to DNA polymerase, elongate this connection fragment by composition using this DNA polymerase, and a flush end-ized double-stranded-DNA fragment is built. And it includes further amplifying this flush end-ized fragment by the approach including contacting this flush end-ized fragment to this DNA polymerase and primer oligodeoxynucleotide (this primer oligodeoxynucleotide containing a longer adapter chain). In said contact, are lower than the melting out temperature of this primer oligodeoxynucleotide from the chain of this flush end-ized fragment complementary to this primer oligodeoxynucleotide. The approach according to claim 6 of performing at temperature higher than the melting out temperature of this short **** of the adapter nucleic acid from this flush end-ized fragment.
12. The approach according to claim 6 of recognition means being one or more restriction endonucleases, and the recognition site being a target sub array, and including further that the process which carries out a probe digests a sample by one or more restriction endonucleases.
13.(a) the nucleic-acid fragment in the sample which generates one or more signals -- identifying -- and -- (b) These fragments are collected. The approach according to claim 12 of containing things further.
14. The approach according to claim 13 the signal generated with the collected fragment does not match the array in a nucleotide sequence database.
The approach according to claim 13 of including further using the part of this fragment which can be hybridized at least as a hybridization probe combined with the nucleic acid which can generate this fragment after digestion by 15.1 or more restriction endonucleases.
16. The approach according to claim 12 of including further removing from a sample the nucleic-acid fragment both which a generation process produces from digestion only at the single end of the nucleic acid which was not digested after this digestion, and this fragment.
17. The method according to claim 16 of making a biotin molecule combine the nucleic acid in a sample by the end before digestion, respectively, and performing this clearance by the approach including making the streptoavidin or avidin which made the nucleic acid in a sample adhere to a solid phase base material contact.
18. The method according to claim 16 of making a hapten molecule combine the nucleic acid in a sample by the end before digestion, respectively, and performing this clearance by the approach including making the anti-hapten antibody which made the nucleic acid in a sample adhere to a solid phase base material contact.
The approach according to claim 12 digestion by 19.1 or more restriction endonucleases leaves a single strand nucleotide lobe to both the digestive end.
20. The approach according to claim 19 the process which carries out a probe includes further making a double strand adapter nucleic acid (each adapter nucleic acid having a complementary end in this lobe generated by the specific thing of one or more restriction endonucleases) hybridize with a digestive sample fragment, using a ligase for the five prime end of the chain of a digestive sample fragment, making the chain of this adapter nucleic acid connect with it, and making a connection nucleic-acid fragment form.

The method according to claim 20 of performing the digestion and connection by 21.1 or more restriction endonucleases in the same reaction medium.
22. Including aforementioned digestion and connection incubating this reaction medium at the 2nd temperature next with the 1st temperature, one or more restriction endonucleases have activity higher than the 2nd temperature at the 1st temperature in that case, and this ligase is the approach according to claim 21 that activity is high, at the 2nd temperature from the 1st temperature.
23. The method according to claim 22 of repeating the incubation in the 1st temperature, and the incubation in the 2nd temperature, and performing them.
24. The approach according to claim 20 the process which carries out a probe includes further removing an end phosphoric-acid radical from DNA in this sample by incubating with the alkaline phosphatase before this digestion.
25. The approach according to claim 24 by which the alkaline phosphatase is unstable with heat, and thermal inactivation is carried out before this digestion.
26. An approach including a generation process amplifying a connection nucleic-acid fragment according to claim 20.
27. It is a method according to claim 26 of making the nucleic acid biosynthesis the aforementioned magnification being performed using nucleic-acid polymerase and a primer nucleic-acid chain, and according [this primer nucleic-acid chain] to this polymerase start.
28. The approach according to claim 27 a primer nucleic-acid chain has 40 - 60% of G+C content.
29. Have Chain with Each Shorter Adapter Nucleic Acid, and Longer Chain, and Longer Chain is Connected with Digestive Sample Fragment. A short chain is dissolved. said generation process -- before a magnification process -- this twist from a connection fragment -- It includes contacting this connection fragment to DNA polymerase, elongating this connection fragment by composition using DNA polymerase, and producing a flush end-ized double-stranded-DNA fragment. a primer nucleic-acid chain -- this twist -- the method according to claim 27 of making only magnification of the flush end-ized double-stranded-DNA fragment with which each different primer nucleic-acid chain is generated after digestion by specific restriction endonuclease including the part which can hybridize the array of a long chain start.
30. Have Chain with Each Shorter Adapter Nucleic Acid, and Longer Chain, and Longer Chain is Connected with Digestive Sample Fragment. A short chain is dissolved. said generation process -- before a magnification process -- this twist from a connection fragment -- It includes contacting this connection fragment to DNA polymerase, elongating this connection fragment by composition using DNA polymerase, and producing a flush end-ized double-stranded-DNA fragment. a primer nucleic-acid chain -- this twist -- the method according to claim 27 of making only magnification of the flush end-ized double-stranded-DNA fragment with which each different primer nucleic-acid chain is generated after digestion by specific restriction endonuclease including the array of a long chain start.
31. magnification -- although a primer nucleic-acid chain is lower than the melting out temperature of this primer nucleic-acid chain from a chain complementary to this primer nucleic-acid chain in process -- this twist from a flush end-ized fragment -- the approach according to claim 30 by which annealing is carried out to a connection nucleic-acid fragment at temperature higher than the melting out temperature of a short adapter chain.
32. The approach according to claim 30 of adjoining the array at the three-dash terminal of a further more long chain, and containing the part of the restriction endonuclease recognition site which remains at the nucleic-acid fragment end after digestion by this restriction endonuclease including a primer with a primer nucleic-acid chain specific to specific restriction endonuclease.
33. The approach according to claim 32 of being that in which the connection nucleic-acid fragment with which each primer specific to specific restriction endonuclease adjoined it by 3' side of the residual part of a restriction endonuclease recognition site, and was amplified by that cause by the three-dash terminal, including one or more nucleotides further adjoins these one or more additional nucleotides, and contains this residual part of a restriction endonuclease recognition site.

34. The approach according to claim 33 which the indicator of the detection of the aforementioned specific primer is made possible, and may be detected identifiable from the primer containing these one or more additional nucleotides from which this primer containing these one or more additional nucleotides of the result specification differs.
35. The approach according to claim 6 a recognition means is the oligomer of the combination of a target sub array, the nucleotide which can be hybridized specifically and a nucleotide false object, or a nucleotide and a nucleotide false object.
36. The approach according to claim 35 by which the nucleic-acid fragment in the sample between the hybridized oligomer is amplified by that cause including a generation process amplifying using the primer containing nucleic-acid polymerase and this oligomer.
37.(a) the nucleic-acid fragment in the sample which generates one or more signals -- identifying -- and -- (b) These fragments are collected. The approach according to claim 36 of containing things further.
38. The approach according to claim 37 the signal generated with the collected fragment does not match the array in a nucleotide sequence database.
39. The approach according to claim 37 of including further using the part of this fragment which can be hybridized at least as a hybridization probe combined with the nucleic acid which can generate this fragment after the magnification using nucleic-acid polymerase and one or more primers.
40. The aforementioned signal is the approach according to claim 1 of including further the display of whether an additional target sub array exists on this nucleic acid in the sample during this existence of a target sub array.
41. The approach according to claim 40 recognized by the approach an additional target sub array includes contacting the oligomer of the nucleic acid in a sample, and the combination of the target sub array of this addition, the nucleotide which can be hybridized and a nucleotide false object, or a nucléotide and a nucleotide false object.
42. A generation process is the approach according to claim 1 of including controlling this signal, when an additional target sub array exists on this nucleic acid in the sample during this existence of a target sub array.
It Includes that Generation Process Amplifies Nucleic Acid in Sample. Additional Target Sub Array 43. Nucleic Acid in Sample, (a) The nucleotide which hybridizes with the target sub array of this addition, and destroys a magnification process, The restriction endonuclease which has the target sub array of this addition as the oligomer of the combination of a nucleotide false object, or a nucleotide and a nucleotide false object, or a (b) recognition site, and digests the nucleic acid in a sample in this recognition site, The approach according to claim 42 recognized by the approach including making it contact.
44. The approach according to claim 12 or 36 of including further that a generation process separates a nucleic-acid fragment with die length.
45. The approach according to claim 44 of including further detecting the nucleic-acid fragment which the generation process separated.
46. The approach according to claim 45 determined from the quantitative level of one or more signals generated with this nucleic acid with which it was determined that the abundance of a nucleic acid including the specific nucleotide sequence in a sample matched this specific nucleotide sequence.
47. The approach according to claim 45 performed by the approach the aforementioned detection includes silver dyeing this fragment, carrying out the indicator of this fragment with DNA insertion coloring matter, or detecting luminescence from the fluorochrome on this fragment.
48. The approach according to claim 45 the display of the die length during existence of a target sub array is the die length of the fragment determined according to said separation and a detection process.
49. The approach according to claim 45 by which said separation is performed using liquid chromatography or a mass spectrometer.
50. The approach according to claim 45 by which said separation is performed using

electrophoresis.
51. The approach according to claim 50 by which electrophoresis is performed in the slab gel using denaturation or a non-denaturalizing medium, or a capillary tube configuration.
52. A target sub array is the approach according to claim 1 of being what generates at least one signal with which one or more predetermined nucleotide sequences in said database are the target arrays, and the target array is not generated by other nucleotide sequences in this database.
53. The approach according to claim 52 the target nucleotide sequence is an array of the generality in this database.
54. the probability for a target sub array to exist in the nucleotide sequence of this database -- about 0.01- the approach according to claim 1 of being about 0.30.
55. A target sub array is the approach according to claim 1 of being what averages although the nucleotide sequence in this database averages and generates the signal generated by none of other nucleotide sequences in this database, and includes existence of a sufficient number of target sub arrays.
56. The number of averages of the signal which the number of pairs of the target sub array which averages and exists in 1 nucleotide sequence in this database is three or more, and is generated from the nucleotide sequence in this database is the approach according to claim 55 of being that from which the mean difference between the die length expressed by the generated signal becomes one or more nucleotides.
57. Target Sub Array is the Following Formula in General. : $\frac{R(R+1)p^2}{2} = A$

It reaches. $\frac{L}{Np^2} = B$

The inside of [type, It averages in the nucleotide sequence from which it differs in an A= this database. number [of the nucleotide sequences from which it differs in an N= this database]; -- average die-length [of the nucleotide sequence from which it differs in an L= this database]; -- number [of R= recognition means]; -- The method according to claim 55 of having the existence probability p given by solving mean difference] between the die length expressed by the signal generated from the array in a number [of the existing target sub array] of pairs;, and B= this database.
58. The approach according to claim 57 A is three or more.
59. The approach according to claim 57 B is one or more.
60. Target Sub Array is the Following Process. : (A) By Simulating Probe Process and Generation Process Which were Applied to Array in Nucleotide Sequence Database The process which determines the array which can generate the pattern of a signal and each signal which may be generated, (b) the process which checks the value of the determined this pattern according to a certain information scale -- and -- (c) Process which chooses a target sub array in order to generate the new pattern which optimizes this information scale Approach according to claim 1 chosen according to the further process to include.
61. The approach according to claim 60 of choosing the target sub array said whose selection process includes the recognition site of one or more restriction endonucleases.
62. The approach according to claim 60 of choosing the target sub array said whose selection process adjoins one or more additional nucleotides, and includes the recognition site of one or more restriction endonucleases.
63. The approach according to claim 60 of being the number of the object arrays which generate at least one signal generated by no nucleotide sequence of an and also [the information scale which one or more predetermined nucleotide sequences which exist in a nucleotide sequence database are the target arrays, and is optimized exists in this database].
64. The approach according to claim 63 of being a large majority of nucleotide sequences to which the target nucleotide sequence exists in this database.
65. The approach according to claim 60 to which said selection process is carried out by

comprehensive retrieval of all the combination of the target sub array of less than about ten die length.
66. The approach according to claim 60 performed by the approach including annealing by which the selection process of a target sub array was simulated.
67. Said Search Procedure is the Following Process. : (A) By Simulating Probe Process and Generation Process Which were Applied to Each Array in Nucleotide Sequence Database the process which determines the array which can generate the pattern of a signal and each signal which may be generated -- and -- (b) In this pattern (i) Die length during existence of the same target sub array as what is expressed by the generated signal, And the same target sub array as what is expressed by the signal which carried out (ii) generation, Or the same target sub array which is expressed by the generated signal and which is the member of a target sub array set, Process which finds out one or more nucleotide sequences in this database that can generate one or more generated this signals by finding the signal containing ****** The approach according to claim 1 of including further.
68. Said Decision Process is the Following Process. : (A) Process Which Searches Existence of Target Sub Array in Nucleotide Sequence of Nucleotide Sequence Database or Target Sub Array Set, (b) The process which finds the die length during existence of this target sub array in the nucleotide sequence of this database or a target sub array set, It reaches. (c) Process in which that a target sub array exists forms the pattern of the signal which may be generated from the array of this found-out database The approach according to claim 60 or 67 of including further.
69. Restriction Endonuclease Generates 5' Lobe at the End of Digestive Fragment. A 2 each chain adapter nucleic acid (a) Consist of the 1st and 2nd continuation part. A shorter nucleic-acid chain (the amount of this part I is a five prime end sub array complementary to the lobe generated by one of the restriction endonucleases), It reaches. (b) Longer nucleic-acid chain which has a complementary three-dash terminal sub array in a part for this part II of a short nucleic-acid chain The approach according to claim 20 of being what is included.
The approach according to claim 69 which has a melting out temperature from a complementary strand with a nucleic-acid chain [shorter than 70.] lower than about 68 degrees C, and does not have an end phosphoric-acid radical.
The approach according to claim 70 a nucleic-acid chain shorter than 71. is the die length of about 12 nucleotides.
The approach according to claim 69 have a melting out temperature from a complementary strand with a nucleic-acid chain [longer than 72.] more expensive than about 68 degrees C, and which array in this database is not complementary, and does not have an end phosphoric-acid radical.
73. The approach according to claim 72 the connected nucleic-acid fragment includes neither of the recognition sites of restriction endonuclease.
The approach according to claim 72 by which thermal inactivation is carried out before 74.1 or more restriction endonucleases' connecting.
The method according to claim 72 of a nucleic-acid chain longer than 75. being the die length of about 24 nucleotides, and having 40 - 60% of G+C content.
76. Restriction Endonuclease Generates 3' Lobe at the End of Digestive Fragment. A 2 each chain adapter nucleic acid (a) Consist of the 1st and 2nd continuation part. A longer nucleic-acid chain (the amount of this part I is a three-dash terminal sub array complementary to the lobe generated by one of the restriction endonucleases), It reaches. (b) Shorter nucleic-acid chain complementary to the three-dash terminal for this part II of a long nucleic-acid chain The approach according to claim 20 of being what is included.
The approach according to claim 79 which has a melting out temperature from a long chain rather than a nucleic-acid chain shorter than 77. is lower than about 68 degrees C, and does not have an end phosphoric-acid radical.
The approach according to claim 77 a nucleic-acid chain shorter than 78. is the die length of 12 base pairs.
The approach according to claim 76 have a melting out temperature from a complementary strand with a nucleic-acid chain [longer than 79.] more expensive than about 68 degrees C, and

it is not complementary, do not have an end phosphoric-acid radical, but the connected nucleic-acid fragment includes [as opposed to / no / the nucleotide sequence in this database] any recognition site of restriction endonuclease.
The method according to claim 79 of a nucleic-acid chain longer than 80. being the die length of 24 base pairs, and having 40 - 60% of G+C content.
81. It is Approach of Identifying or Classifying Nucleic Acid, and is the Following Process. : (A) Process Which Carries Out Probe of the Nucleic Acid with Two or More Recognition Means, However, in order that each recognition means may generate an one-set signal, a target nucleotide sub array or an one-set target nucleotide sub array is recognized. And as for each signal, one of this target sub array or the target sub arrays of this set indicates whether exist in this nucleic acid. It reaches. (b) The nucleotide sequence database containing many known nucleotide sequences of the nucleic acid which may exist in a sample The process searched about the array which matches the signal of the generated set, However, [whether the same target sub array as what is displayed that the array from this database exists with the signal of the set which carried out (i) generation is included, and] Or the target sub array which is the member of the target sub array set displayed to exist is included. And the target sub array displayed not to exist with the signal of the set which carried out (ii) generation, Or when it does not include the target sub array which is the member of the target sub array set displayed not to exist, the array from this database matches the signal of this set, and this identifies or classifies this nucleic acid. The above-mentioned approach of coming to contain each process.
82. The approach according to claim 81 displayed by the hash code whose signal of said set is a binary digit.
83. The approach according to claim 81 said process which carries out a probe brings about the quantitative signal of the number of existence of the target sub array in this nucleic acid, or the number of members of this target sub array set.
84. It is the approach according to claim 83 which matches the signal of the set which generated this array when the array from said database did not include the target sub array in this target sub array set that is displayed to include the same target sub array during this array, and not to exist with the same number of existence as the thing in a quantitative signal, and that is displayed not to target-factice-arrange or not to exist.
85. The approach according to claim 81 many nucleic acids are DNA.
86. The approach according to claim 85 of a recognition means being the oligomer the indicator of the detection of the combination of a nucleotide, a nucleotide false object, or a nucleotide and a nucleotide false object of was made possible, and including that said process which carries out a probe makes this nucleic acid hybridize with this oligomer.
87. The approach according to claim 86 which the approach the oligomer the indicator of the detection of was made possible includes detecting luminescence from the fluorochrome indicator on this oligomer, or this indicator oligomer is arranged, and light is scattered from an optical pipeline, and is detected by the approach including detecting the dispersion.
88. The approach according to claim 86 a recognition means is the oligomer of a peptide-nucleic acid.
89. The approach according to claim 86 a recognition means is the set of a DNA oligomer, the DNA oligomer containing a universal nucleotide, or the DNA oligomer that degenerated in part.
90. Search Procedure is the Following Process. : (A) By Simulating Probe Process Applied to Each Array in Nucleotide Sequence Database The pattern of the set of the signal of existence of this target sub array or this target sub array set or nonexistence, and the process which determines the array which can generate the signal of each set by this pattern -- and -- (b) By finding the generated set and the set which matches in this pattern The process which finds out one or more nucleotide sequences which can generate the signal of the generated set, However, it is displayed that the target sub array which is the member of the same target sub array set as what is displayed that the set of the signal from this pattern exists with the signal of the set which carried out (i) generation displayed to target-factice-arrange or to exist exists. And when it is displayed that the target sub array which is the member of the target sub array set which is displayed not to exist with the signal of the set which carried out (ii) generation, and which is

displayed not to target-factice-arrange or not to exist does not exist, The one-set signal from this pattern matches the signal of the generated set. The approach according to claim 85 of including further.

91. Target Sub Array is the Following Process. : (A) By Simulating Probe Process Applied to Each Array in Nucleotide Sequence Database (i) Display a target sub array, or existence of this target sub array set or nonexistence.

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]
They are identification, a classification or the approach of carrying out a quantum, and equipment about the DNA array in a sample without performing sequencing. This application is the United States patent application 08th of application / continuation-in-part of No. 547,214 (attached on October 24, 1995) on the same day, and quotes all of those contents here.
the acknowledgement number 70th according [this invention] to National Institute of Standards and Technology -- it was carried out to the basis of assistance of the U.S. Government as NANB5H1036 No. The U.S. Government has a part of right about this invention.
1. Field of invention The fields of this invention are the quantum of a DNA array, identification or a judgment, and a classification. When specified more, sequencing is carried out in any way, and is twisted and they are the comparison of the all DNA array in a sample or the quantitive classification of a gene, and a manifestation, or identification preferably.
2. Background In the past ten years, our knowledge on the basis of the molecule of a life had become increasingly clear [that a spatial manifestation determines the progress in progress i.e., the health, and the illness of all lives] in time [a gene] in connection with having changed revolutionarily by biology and genome research. Science progressed to the recognition of the importance of the interaction of the deletion of the multiplex gene by the environmental factor in the onset factor of the more complicated disease of large majorities, such as cancer, from an understanding of what the deletion of only one gene causes a failure recognized as heredity traditionally [thalassemia etc.]. In the case of cancer, it is proving that the present scientific proof has a role of a cause by which the manifestation and the multiplex deletion by which the gene used as some shafts was changed serve as KII. It has an onset factor with the same said of other complicated illnesses. In this way, the correlation established between gene expression, health, or an illness condition is more perfect, and if it becomes that reliable, the illness will be recognized, diagnosed and treated better.
Since this important correlation is established according to the quantum measurement and the classification of a DNA manifestation in an organization sample, the quick and economical approach for it has a large meaning. A genomic DNA ("gDNA") array is a DNA array which occurs in the nature which constitutes the genome of a cell. By any cases, a gene or the condition of gDNA, and a manifestation are determined by the presentation of the total cell messenger RNA ("mRNA"), and this is compounded by the modulatory imprint of gDNA. A complementary DNA ("cDNA") array is compounded by reverse transcription from mRNA. Although there is cDNA from the total cell mRNA about then, it also determines the manifestation of gDNA in the cell of fixed time amount. After all, quick and economical detection of all DNA arrays especially cDNA, or gDNA is desired. Especially these detection is quick, and it is exact, and it is desirable if quantitive.
Conventionally, the specific DNA analytical skill of a gene was demanding sequencing of until to some extent without targeting the decision or the classification of all substantial DNA which displays the total cell mRNA in a DNA sample. Analytical skill has aimed general at determining and analyzing one, or two known or strange gene sequences about 1 time when cDNA and gDNA exist. With these techniques, the probe compounded in order to recognize specifically only one a

specific DNA array or a specific gene by hybridization has been used. (For example, refer to Watson et al., 1992, Recombinant DNA, chap 7, W.H.Freeman, and New York.) Application of these approaches to the problem of recognizing all the arrays in a sample becomes complicated and noneconomic further.

The one present approach for discovering and carrying out sequencing of the strange gene departs from an array-ized (arrayed) cDNA library. mRNA is isolated from a specific organization or a specific sample, it clones in a suitable vector, and this is put into a plate by the approach that the descendant of each vector which produced the clone of one cDNA array next can be separated and identified. Then, the probe of such a duplicate object of one plate is carried out by the indicator DNA oligomer chosen so that it might hybridize with cDNA which displays the target gene in many cases. The colony which produces the target cDNA is discovered and isolated by this, and sequencing of the cDNA is collected and carried out. Then, sequencing is carried out by applying the dideoxy chain ending method (Sanger et al., 1977, "DNA sequcnsing with chain terminating inhibitors", Proc.Natl.Acad.Sci.USA 74 (12):5463-5467) of Sanger to this isolated insert.

The DNA oligomer probe for the strange gene used for colony selection is compounded so that it may hybridize only with cDNA to the target gene preferably. As one approach for attaining this singularity, there are some which leave the protein product of the target gene. If the partial array of the PEPECHIDO fragment of 5-10 **** can be determined from the active region of this protein, the degeneracy oligonucleotides of corresponding 15-20 **** which carries out the code of this peptide are compoundable. Typically, the aggregate of this degeneracy oligonucleotide is enough to identify only a corresponding gene. Similarly, in order to produce a single radioactive probe, it can be used for any information which draws the nucleotide sub array of the die length of 15-30.

Also in another present approach for searching for the known array in cDNA prepared from the organization sample, or gDNA, a complementary single gene or a single array probe is used for the peculiar sub array of the already known gene sequence. For example, the manifestation of the specific oncogene in a sample can be determined by carrying out the probe of the cDNA of the organization origin using the probe originating in 1 sub array of the array indicator of the discovered oncogene. Similarly, rare ******(ing) again, such as TB bacillus or HIV, can determine these by carrying out the probe of the gDNA to one gene of this pathogen with a specific hybridization probe about existence of a difficult pathogen. This can be determined by carrying out a probe only to the variation allele on appearance using a complementary allele specific probe about existence of the heterojunction of the variation allele in a normal individual, or existence of the gay junction in an embryo (for example, Guo et al., 1994, Nucleic Acid Rcsearch, 22:5456 to 65 reference).

When applying to determining all the genes discovered in the given organization sample, the above-mentioned example is typical. All the present approaches of using a single radioactive probe need the separate probe of thousands to tens of thousands. It is presumed that one human cell can discover the gene of about 15,000 to 15,000 to coincidence typically, and the most complicated organization, for example, a brain, can discover even the one half of a human genome (Liang et al., 1992, "Differential Display of Eukaryotic Messenger RNA by Meansof the Polymerase Chain Reaction, Science, 257:967-971). Thus, these application that needs many probes is too complicated clearly, is not economical, and practical, either.

The another present approach of a class learned as sequencing ("SBH") by hybridization uses the combination of the probe which is not specific for a gene in contrast with this (Drmanac et al., 1993, Science 260:1649-52; U.S.Patent No.5, 202 and 231, Apr 13, 1993, to Drmanac et al.). It is necessary to carry out the probe of the one single cDNA clone in the typical example of activation of SBH for determining a strange gene by fixed die length, for example, all the oligomer of DNA that consists of a hexamer altogether. One set of all the oligomer of the fixed die length compounded without the room of such selection is called a combination probe library. The partial DNA array about cDNA is reconstructible with algorithm actuation from the knowledge of the result of all the hybridization about one combination library, for example, the result of all hexamer BUROBU of 4096. Since the sub array at least repeated cannot be determined completely, a full

array cannot be determined. SBH applied to the classification of a known gene is called an oligomer array signature (signature) ("OSS") (Lennon et al., 1991, Trends In Genetics 7(10):314-317). A single clone is classified according to this technique on the basis of the pattern of the hit of the probe to all combination libraries or a characteristic sublibrary. It is required for an organization sample library to be arranged by this in a clone, and for each clone to include only one pure array from a library. This is inapplicable to mixture.
All of these typical present approaches aim at discovering one array in 1 array of the clone to which each discovers one array from 1 organization sample. These do not aim at the quickness of all the DNA arrays in mixture of an array, such as the specific total cell cDNA or a gDNA sample, and economical, quantitive, and exact property decision.
These application in such a technical problem cannot be performed. Even when DNA of one clone by sequencing, furthermore the decision with thousands of arrays of all samples are quick for an economical and useful diagnosis, they are not cheap, either. the inside of thousands of probes or combination library where the technique of the gene decision on the basis of the present probe or a classification is specific in each possible gene by which a gene should be observed known or irrespective of whether to be strange -- at least -- thousands -- or tens of thousands of probes are needed further. Furthermore, a sample needs to be array-ized in the clone to which, as for all of these approaches, each discovers the single gene of a sample.
Another present technique known as a DIFERENSHARU display (differential display) aims contrastive with the typical present gene decision and the classification technique which were indicated until now at carrying out discernment judgment (fingerprint) of the mixture of the gene discovered [what / is seen to the pooled cDNA library] again. However, it asks for deciding whether this discernment judgment only has two same samples or it differs. The attempt which determines quantitatively the gene expression as which specification was determined qualitatively is not performed (Liang et al., 1995, Current Opinions in Immunology 7:274-280; Liang et al., 1992, Science 257:967-71; Welsh et al., 1992, Nucleic Acid Res.20:4965- 70; McClelland et al., 1993, Exs 67:103- 15; Lisitsyn, 1993, and Science 259:946-50). A DIFERENSHARU display makes the DNA sub array of various kinds of die length amplify using polymerase chain reaction ("PCR"), and is decided by placing between the hybridization parts of the primer which chose these as arbitration. Ideally, the pattern of die length observed is characteristic of the histogen by which the library was prepared. One primer used on a DIFERENSHARU display is oligo (dT) typically, and another is the oligonucleotide of one or more designs [which were hybridized in the 200-300 base pair of the Polly da tail of 1cDNA in a library] arbitration. By this, the fragment with which the die length from electrophoresis separation to 200 -300 base pair was amplified should generate the band which is characteristic of a sample and can be identified. Change of the gene expression of an organization is observable as change of one or more bands.
Although the pattern of characteristic band formation is developed, the attempt which links these patterns with specific gene expression is not performed. A specific gene cannot be reached in the primer of the 2nd arbitration. To the 1st, an PCR process is not ideally specific. The mismatch ("bubbles") of 1 to 2 and 3 base pair ("bp") is permitted by the annealing step of a low tautness used typically, and there is tolerance [which starts a new chain by Taq polymerase / sufficient]. As one information for identifying all manifestation genes in the location of a single sub array, or its nonexistence, it is inadequate for the 2nd. The information on the die length from the primer of arbitration to the Polly dA tail is not accepted [3rd] to be a characteristic thing to the array which generally exists by the versatility of processing of 3' untranslation region of a gene, the versatility of a polyadenylation process, and the variability of the priming to the reiterative sequence in an exact location. In this way, even if a band is formed, in many cases, it will become indistinct by nonspecific existence of a background array. The bias used as the high G+C content of known PCR and a short array also limits the singularity of this approach further. In this way, generally this technique is limited to the "discernment appraisal" sample about identity or a nonidentity nature judging, and is eliminated from the use in the quantitive decision of the manifestation from which the gene which can be identified differs.
While the component of the cDNA mixture prepared from the organization sample of the present

approach for the classification of a gene or a DNA array or decision is quick and it is economical, it is necessary to improve the capacity to carry out a quantitive and specific judgment. Generalization of the background mentioned above clarifies the defect of some typical present approaches.

3. Epitome of invention One purpose of this invention is offering the approach of carrying out quickness, an economical, quantitive, and exact judgment, or a classification without actually carrying out sequencing of the DNA for one of the genomes or complementary DNA arrays of mixture of arrays, such as a thing, which can be guided from the array or organization sample of a DNA array, especially a single array clone. The defect of the background technique clarified here is solved by this. This purpose is realized by there being the special feature from the DNA array in the sample to analyze, and making two or more detectable signals generate. Preferably, it classifies according to an individual according to the specific signal with which it generates each specific DNA array in a sample, and in order to determine according to an individual with reference to the database of the DNA array which may exist in a sample, all of signals are summarized and it has epicritic [sufficient] and decision nature. It depends for the reinforcement of the signal used as the index of a specific DNA array on the abundance of the DNA quantitatively. Or the main fractions of a DNA array can be slightly classified according to the combination of a signal into two or more sets of the separate array of about 2 to 4.

Still more nearly another purposes are the fewest possible recognition reaction of a number, and making many signals generate slightly about five to 400 **** preferably from the measurement as a result of about 20 to 50 reaction. A quick and economical judgment is not attained in what each DNA array in the sample containing complicated mixture requires the separate reaction in a peculiar probe as. Preferably, or it is a majority of signals which each recognition reaction can identify, a characteristic pattern is generated and it is proportional to the amount of the specific DNA array in which this exists quantitatively. Furthermore, a signal is detected and measured by the minimum desirable observation, and these can carry out to coincidence preferably.

A signal is desirable, is optical, is generated by the fluorochrome indicator, and is detected by the automation optical detection technique. These approaches are used, and many each indicator partial numerators are discriminable even if these are in the same filter spot or a gel band. By this, detection of the signal generated in the reaction and coincidence of multiplicity is attained. Or this invention can be easily fitted to another indicator system, for example, argentation gel. Even when it is optical especially, or since the system of the arbitration which detects a single molecule also with the technique of others, such as a scan or a tunneling microscope, improves a quantitive property remarkably, it becomes very advantageous because of use of this invention.

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DRAWINGS

[Drawing 1]

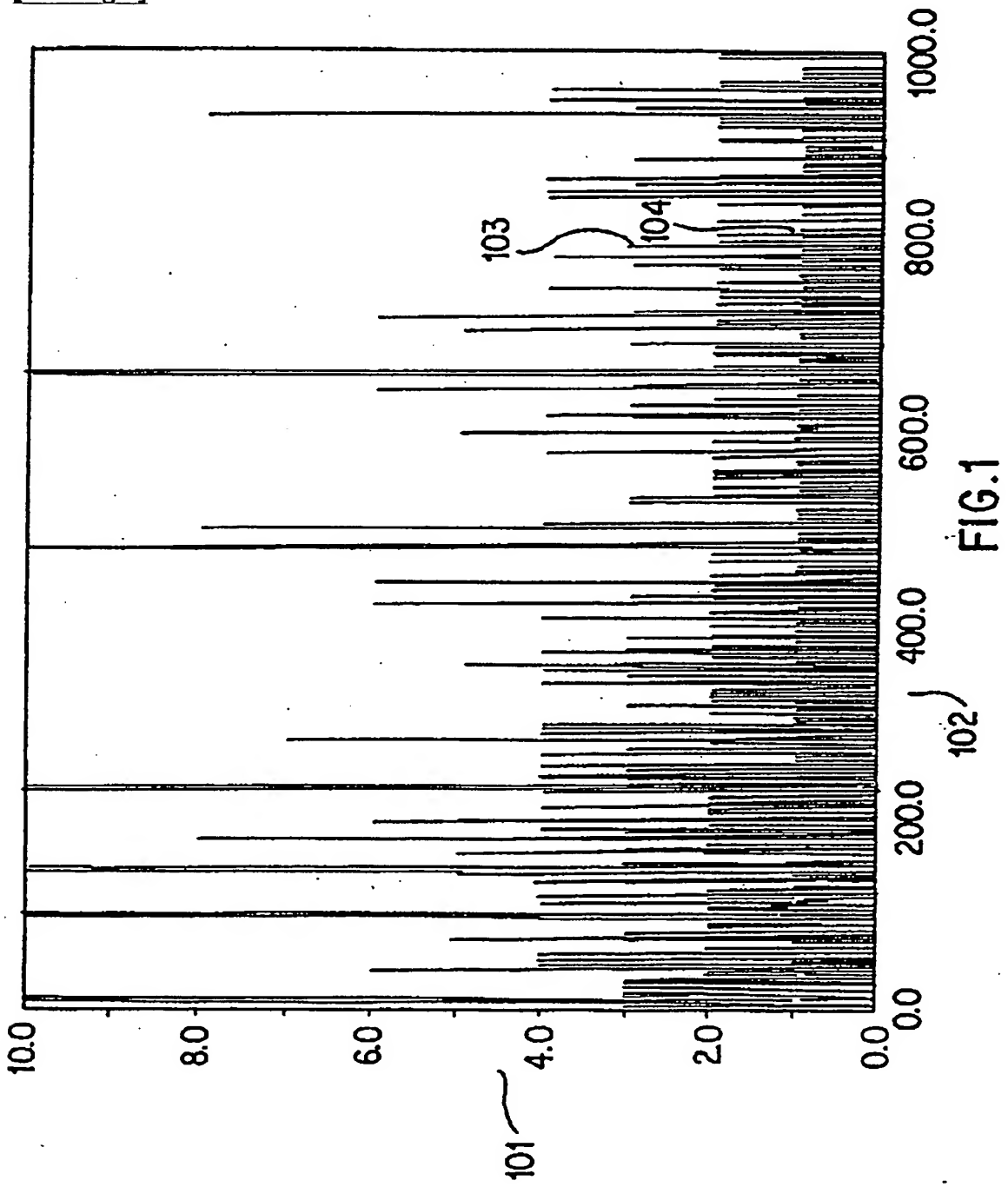


FIG.1

[Drawing 2]

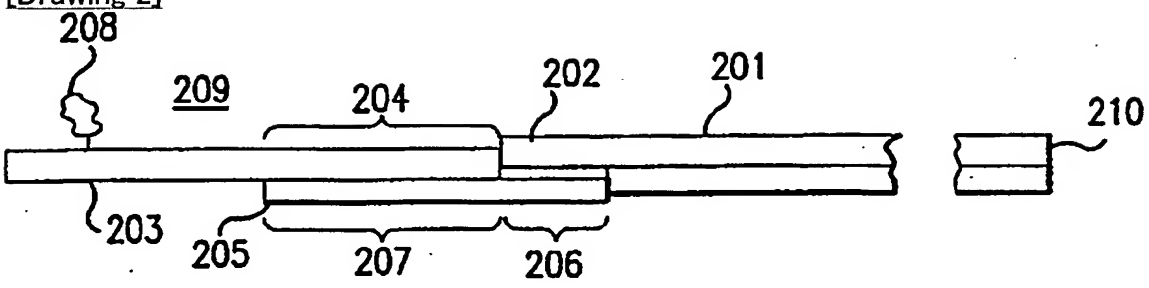


FIG.2A

5'-AGC ACT CTC CAG CCT CTC ACC GAA-3' (配列番号：1)

250 3'- AG TGG CTT CTAG-5' (配列番号：7)

5'-ACC GAC GTC GAC TAT CCA TGA AGC-3' (配列番号：42)

251 3'-GT ACT TCG TCGA-5' (配列番号：44)

FIG.2B

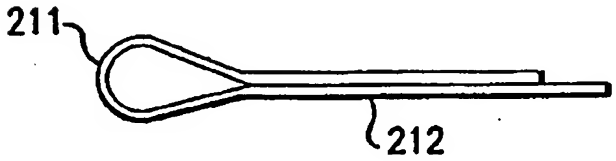


FIG.2C

[Drawing 2]

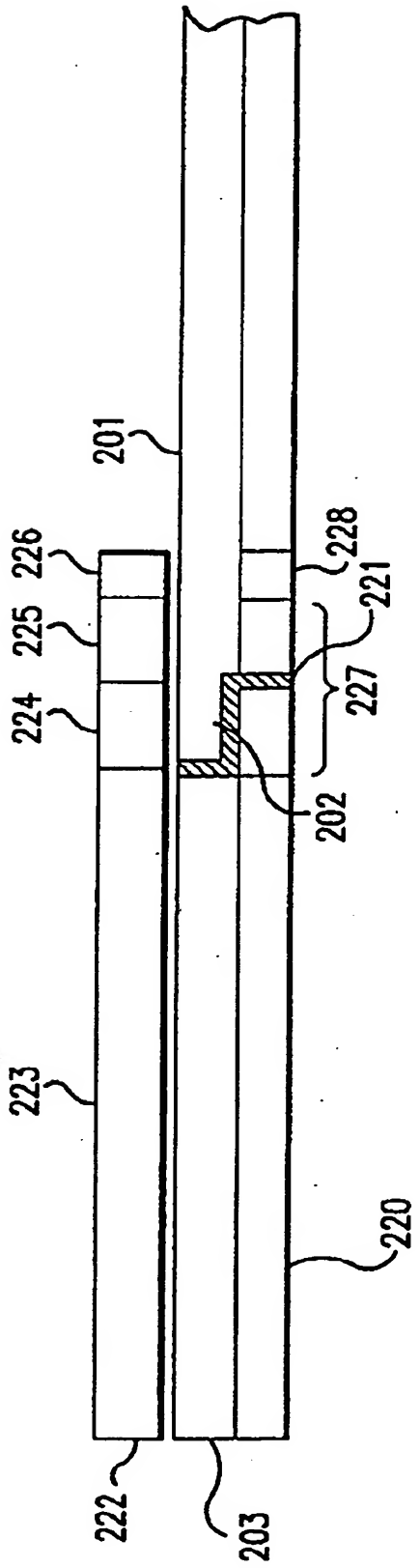


FIG.2D

[Drawing 3]

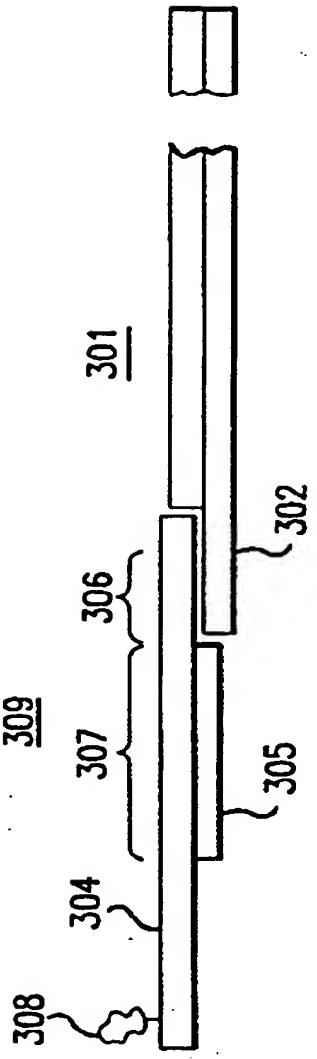


FIG.3A

5'-AGCACTCTCCAGCCTCTCACCGAGCATG (配列番号：55)
3'-AGTGGCTC

FIG.3B

[Drawing 4]

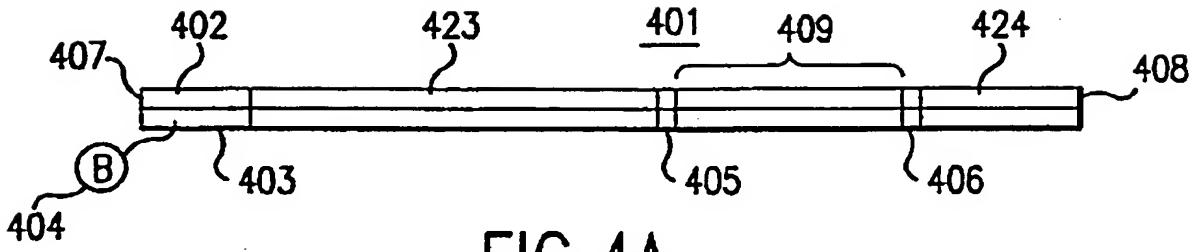


FIG.4A

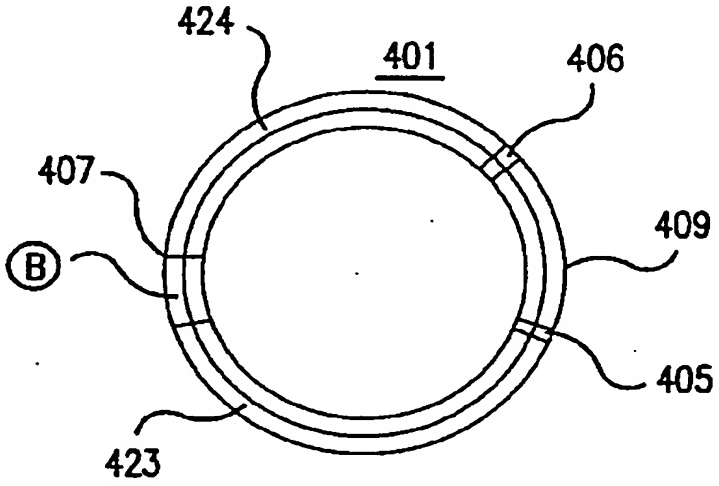


FIG.4B

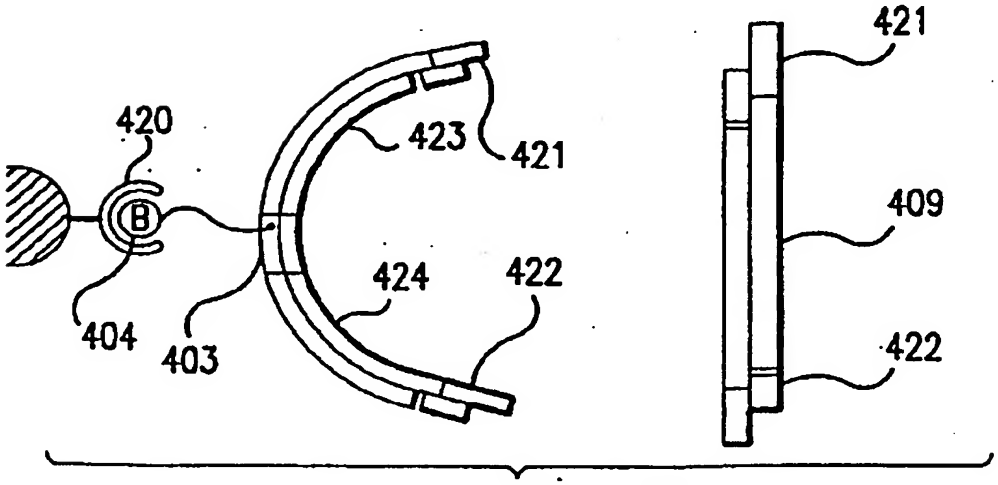


FIG.4C

[Drawing 5]

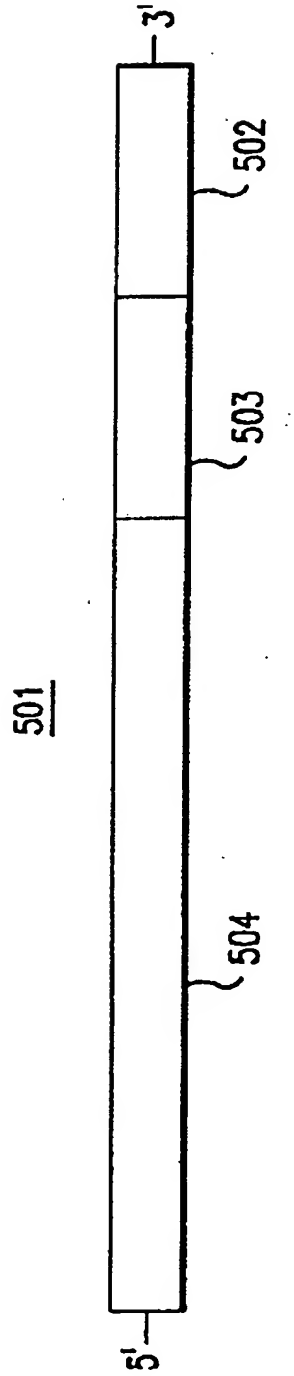


FIG.5

[Drawing 6]

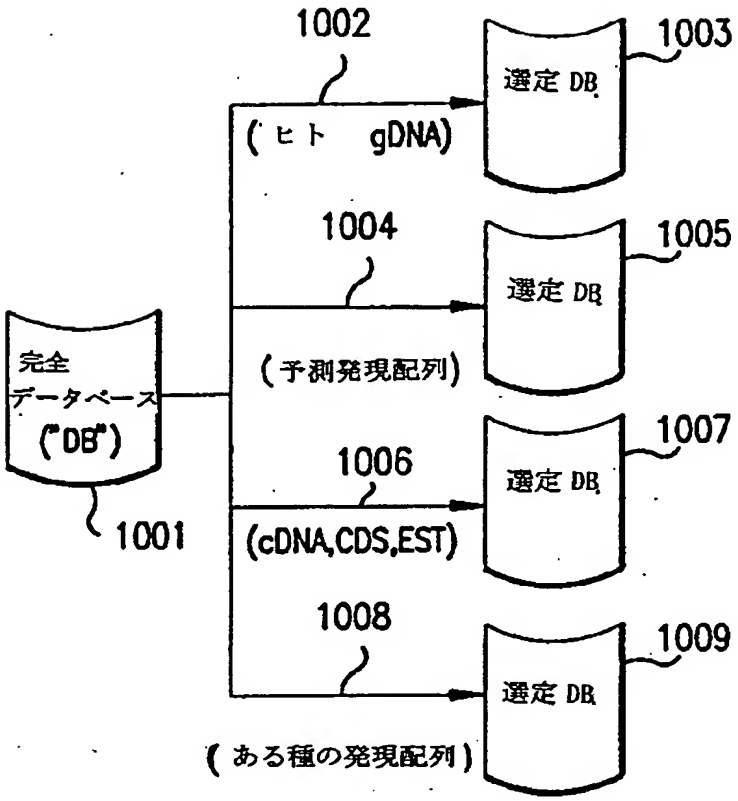


FIG.6A

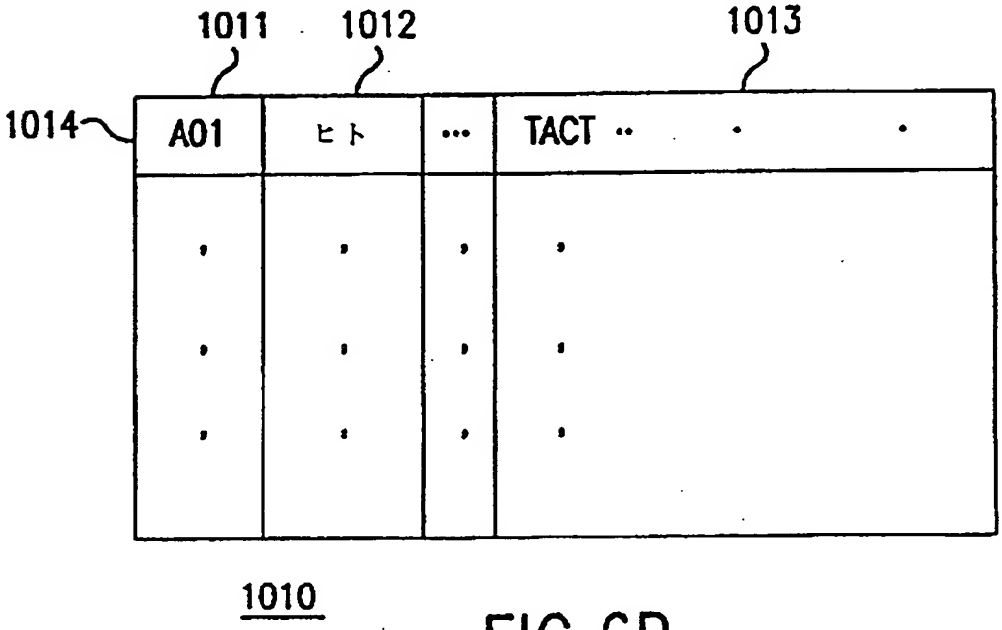


FIG.6B

[Drawing 7]

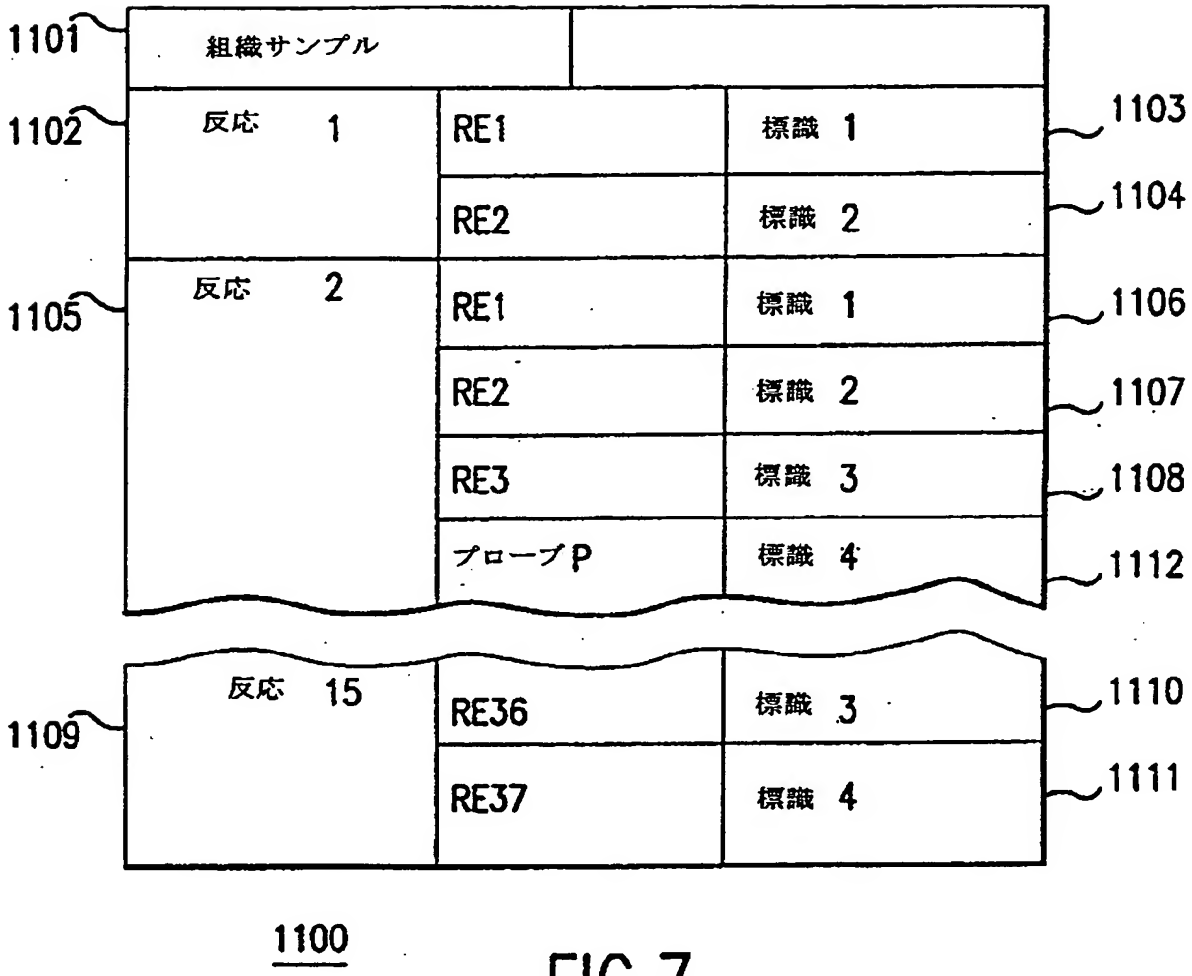


FIG.7

[Drawing 8]

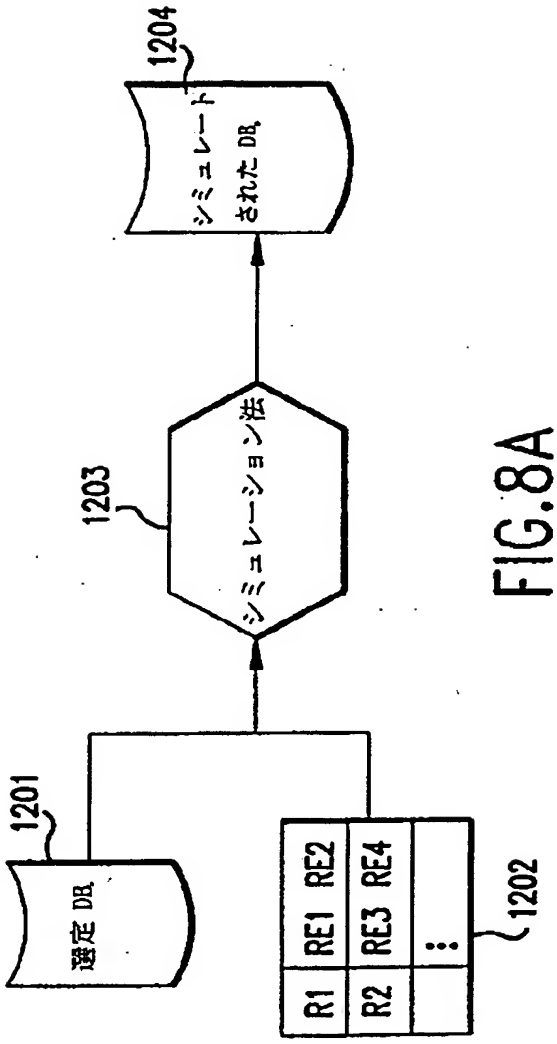


FIG.8A

[Drawing 8]

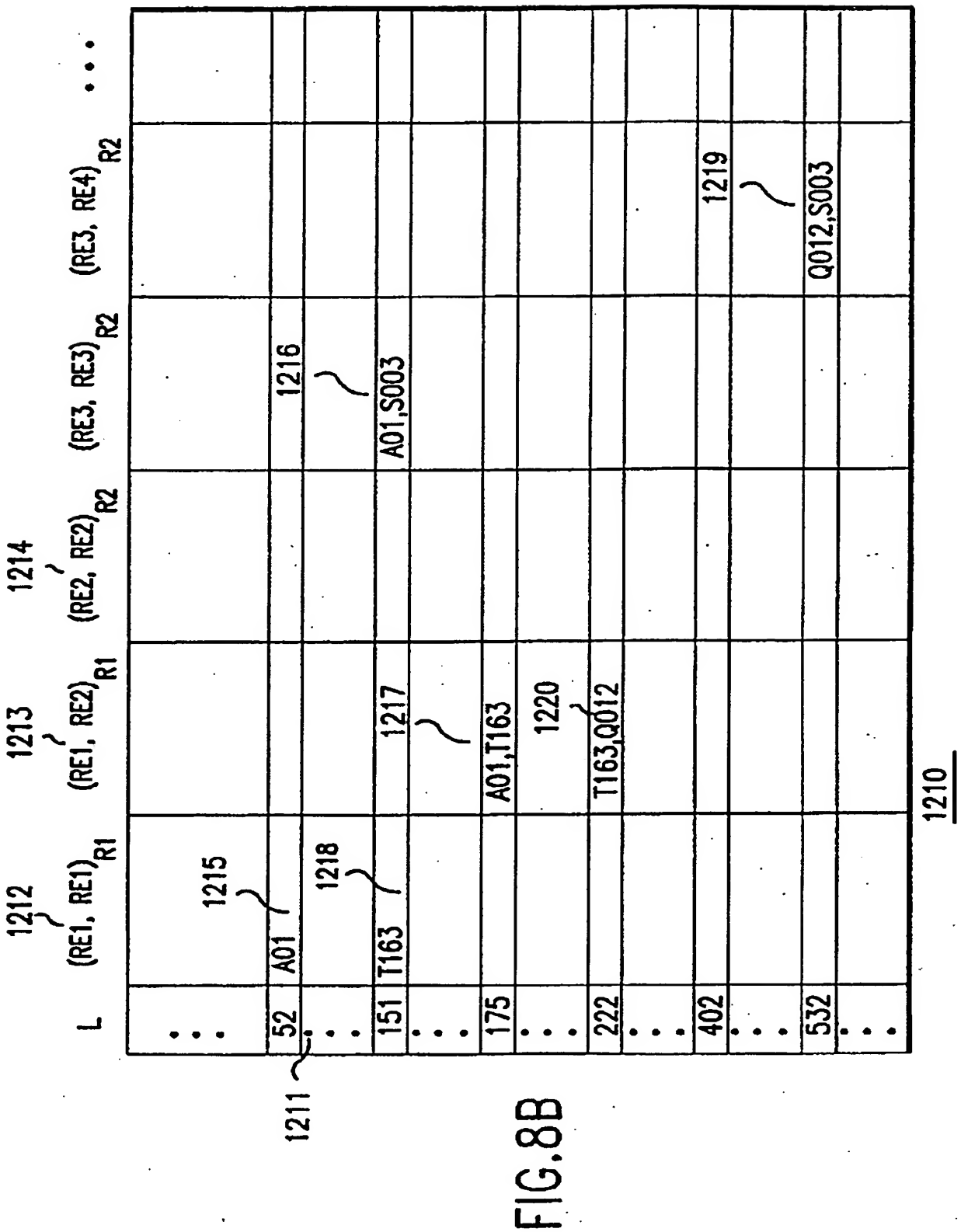


FIG.8B

[Drawing 9]

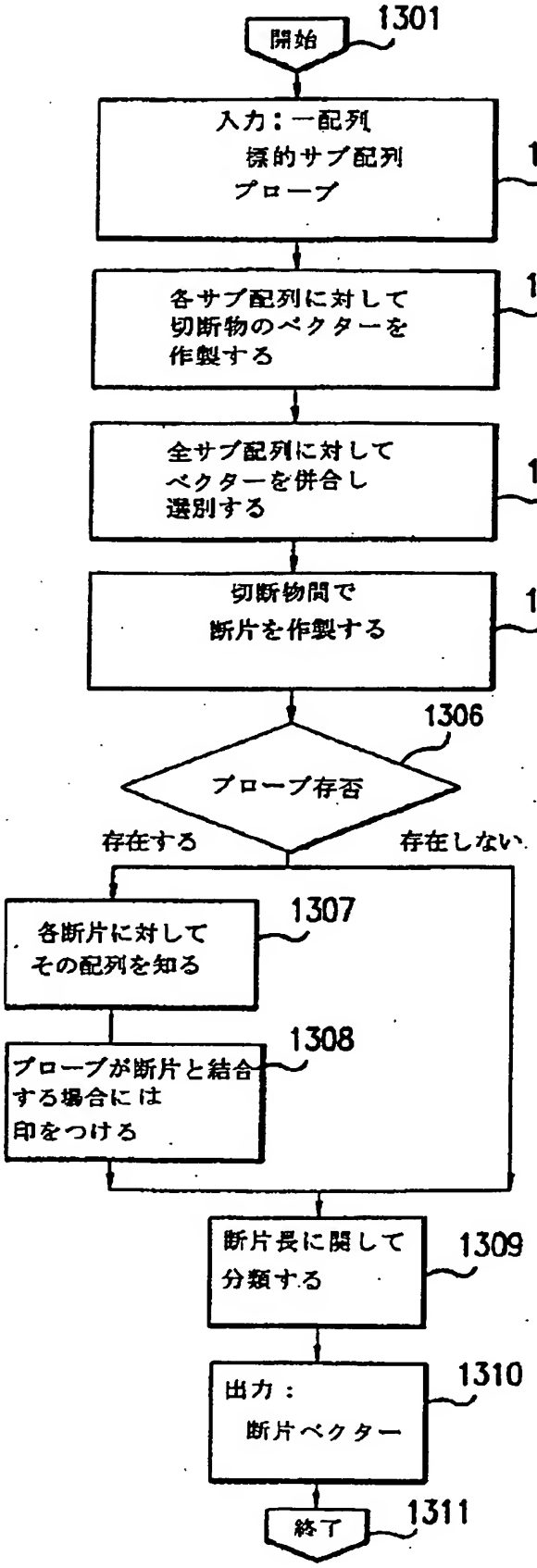


FIG.9

[Drawing 10]

1403

1401 V1 = ([10,14]$_{RE1}$,[62,66]$_{RE1}$,[610,614]$_{RE1}$)

1402 V2 = ([237,241]$_{RE2}$,[388,392]$_{RE2}$)

切断物のベクター

FIG.10A

1404 1405 1406 1407

([10,14]$_{RE1}$,[62,66]$_{RE1}$,[237,241]$_{RE2}$,

[388,392]$_{RE2}$, [610,614]$_{RE1}$)

断片物の併合及び選別ベクター

FIG.10B

1408 1409

[RE1,RE1,*,52] [RE1,RE2,*,175] [RE2,RE2,*,151]

[RE1,RE2,*,222]

断片

FIG.10C

1410

[10,62] [62,237] [237,388] [388,610]

断片配列

FIG.10D

[Drawing 10]

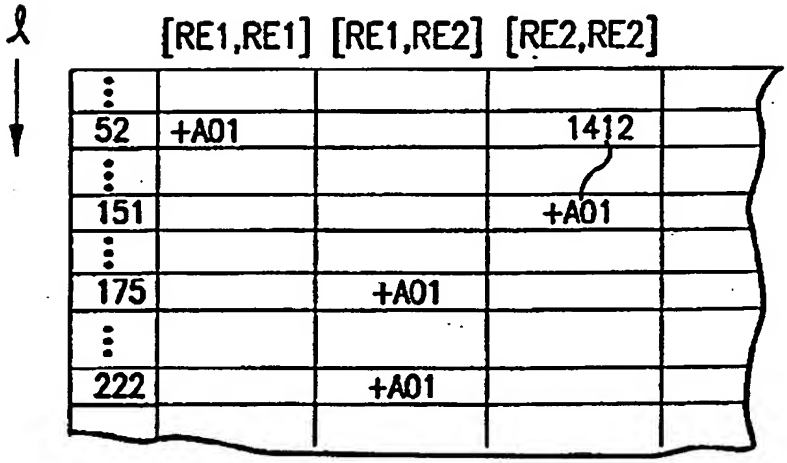
1411

([RE1,RE1,N,52] , [RE2,RE2,y,151] ,
[RE1,RE2,N,175] , [RE1,RE2,N,222])

1412

選別断片ベクター

FIG.10E



ダイジェスト表

FIG.10F

[Drawing 11]

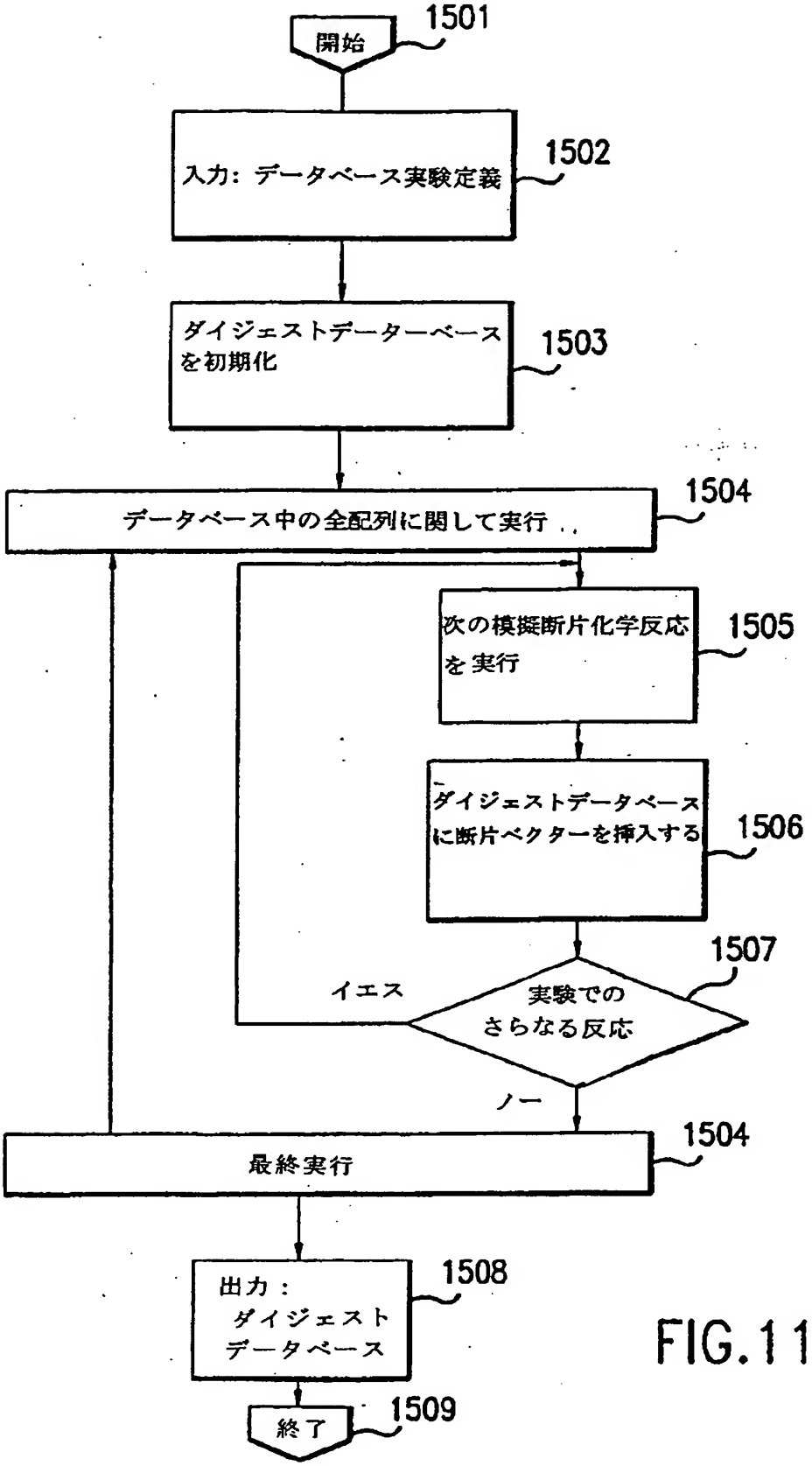


FIG.11

[Drawing 12]

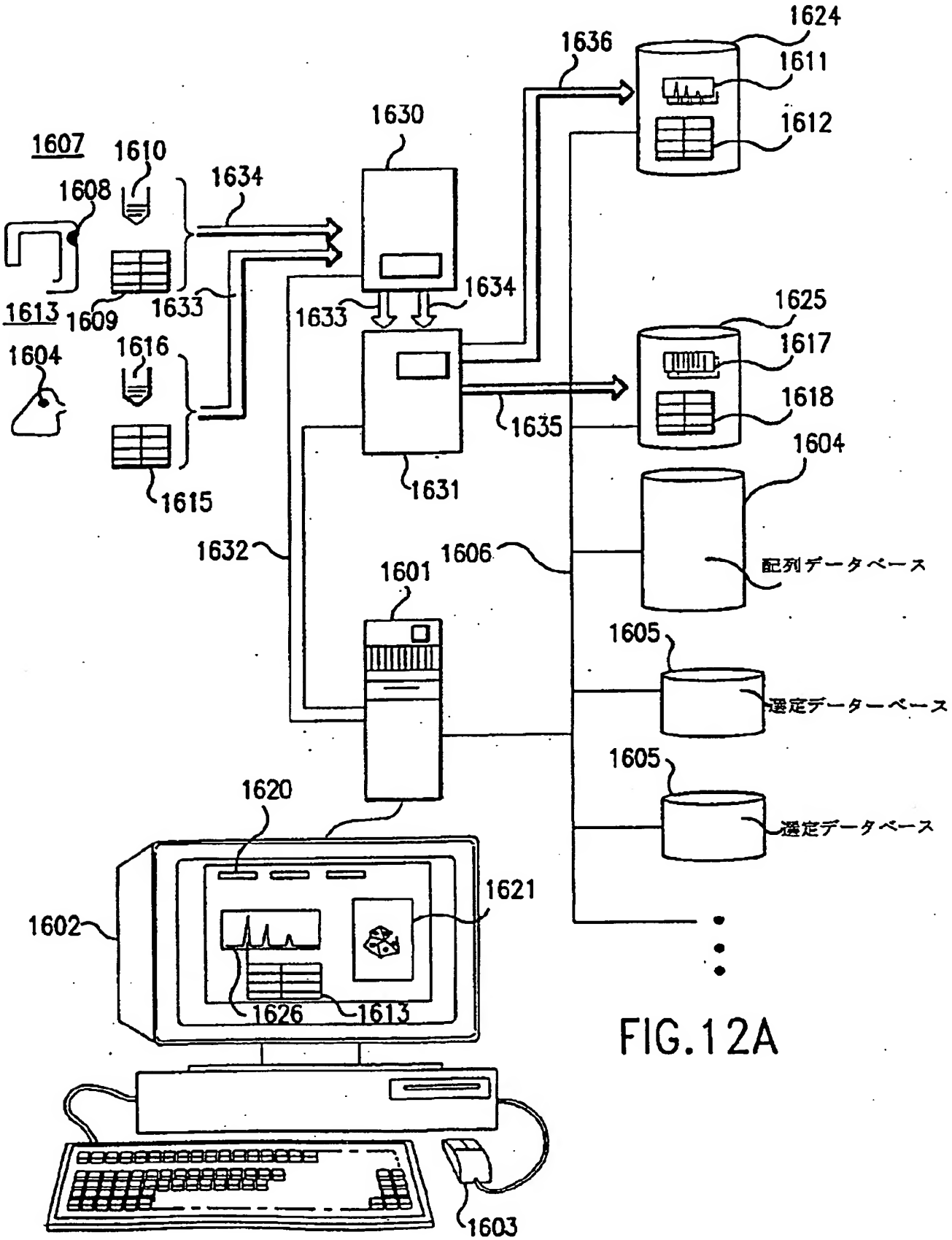


FIG.12A

[Drawing 12]

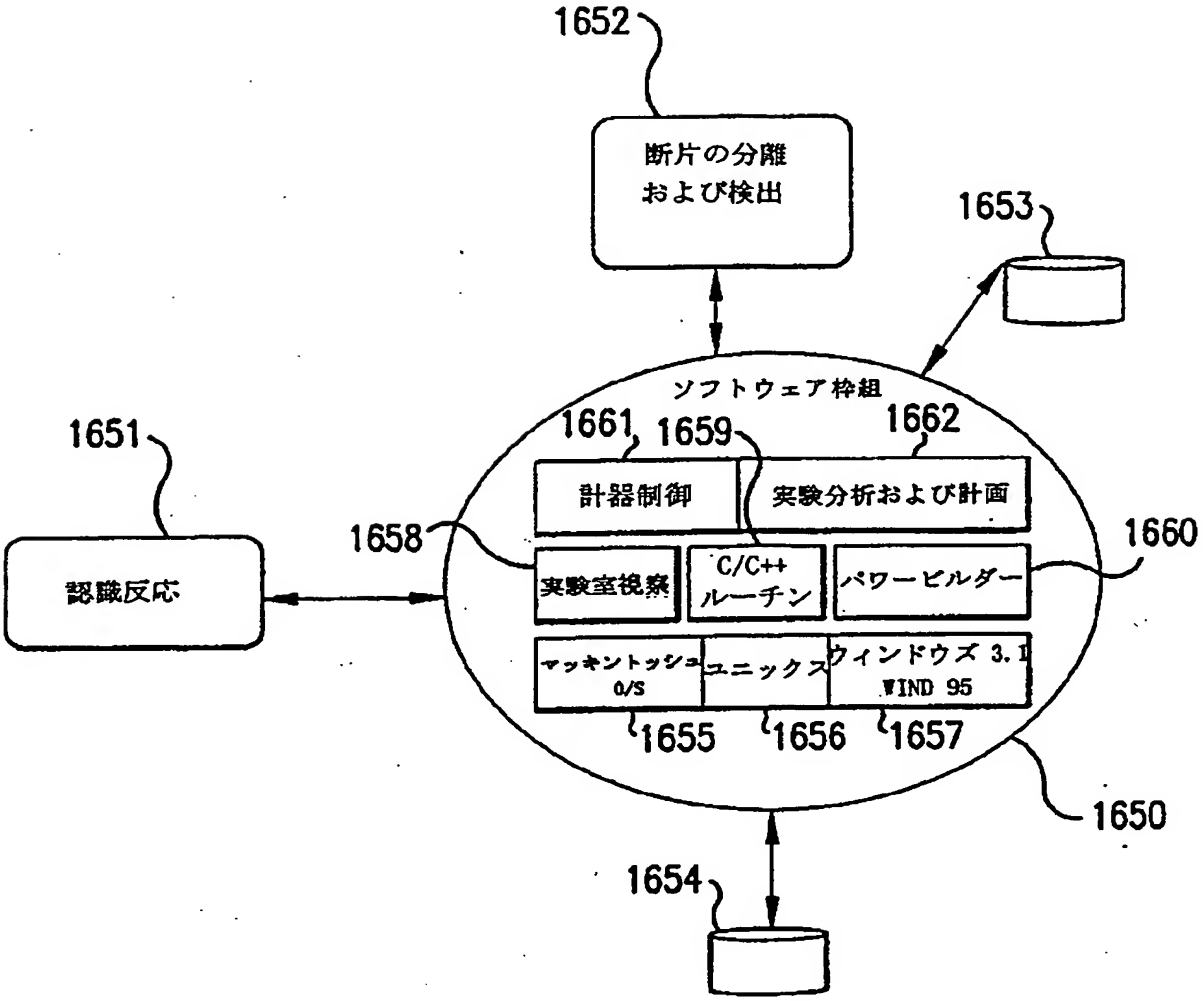


FIG.12B

[Drawing 12]

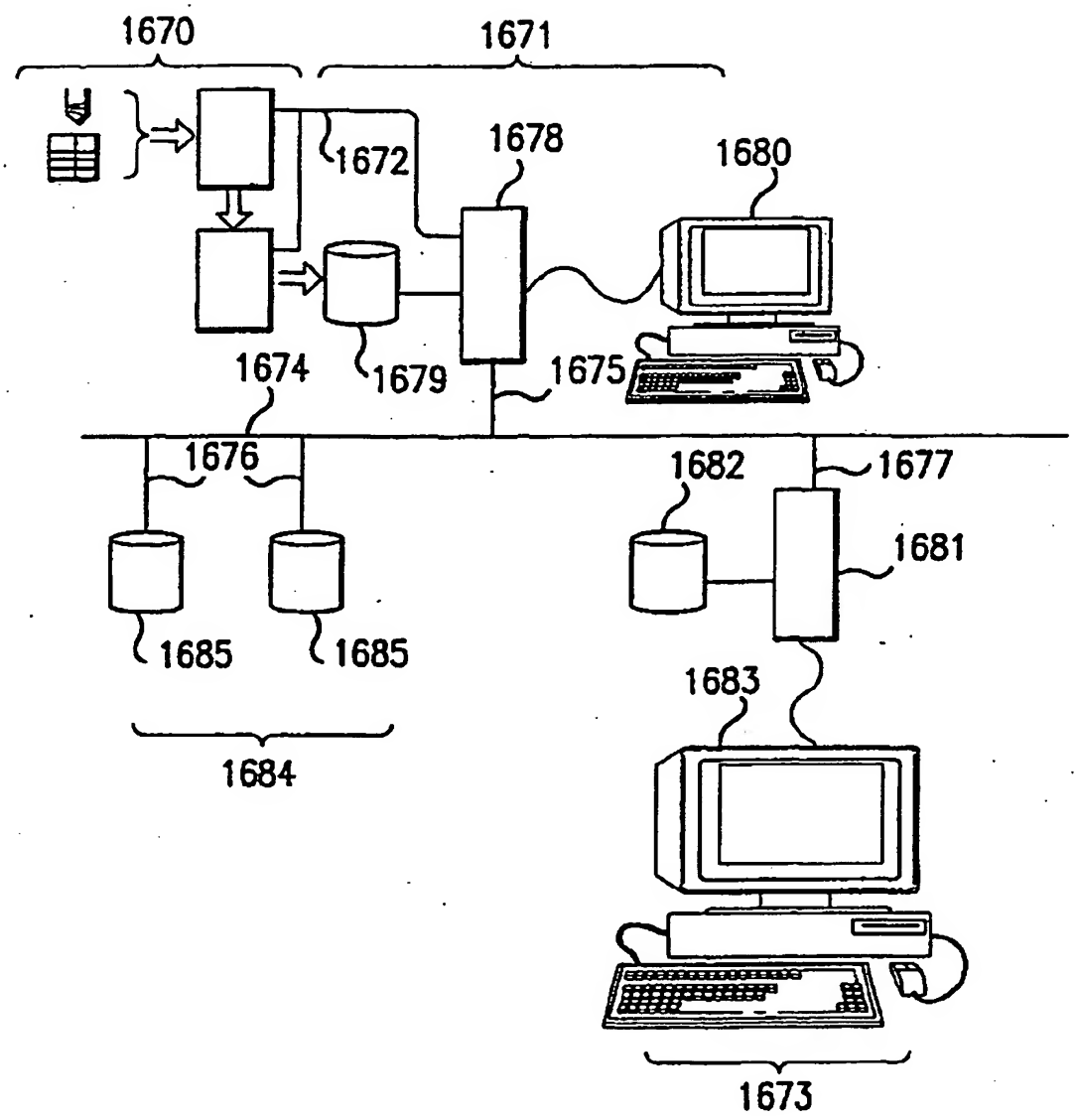


FIG.12C

[Drawing 13]

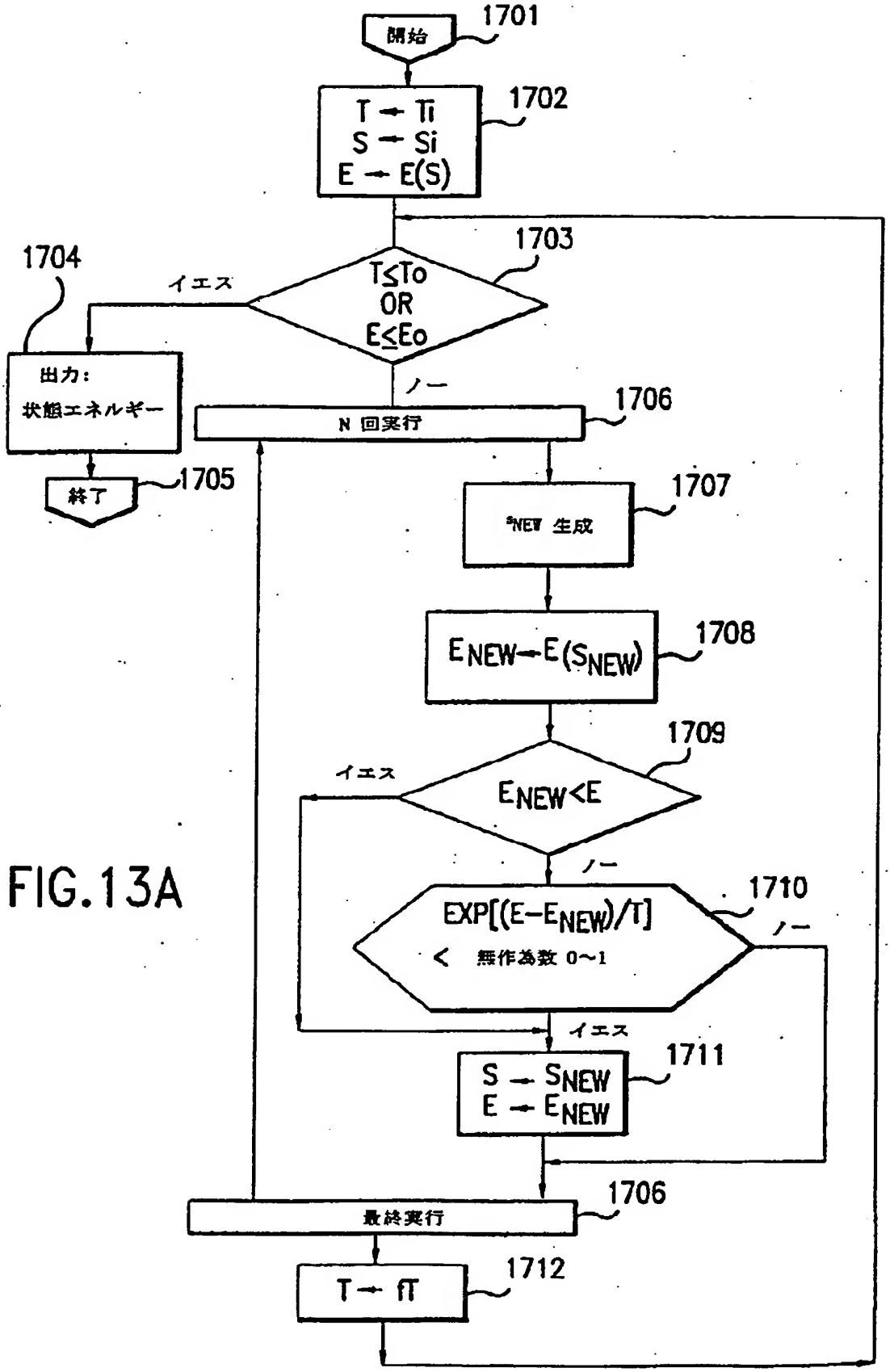


FIG.13A

[Drawing 13]

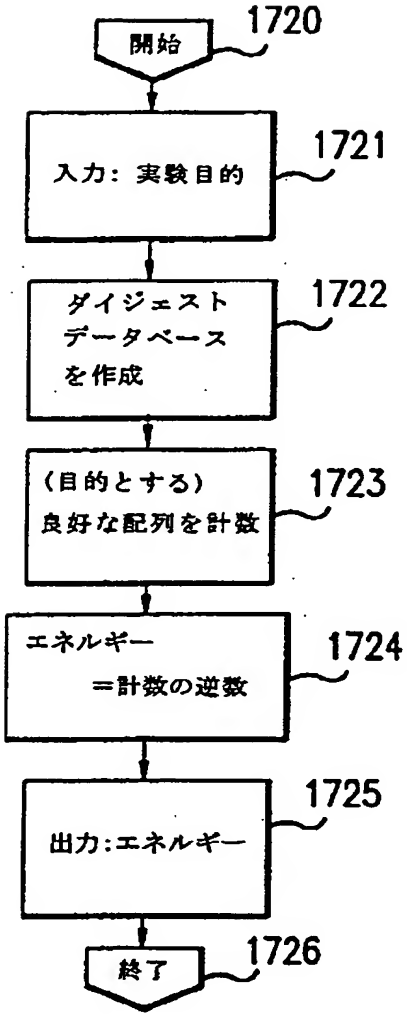


FIG.13B

[Drawing 14]

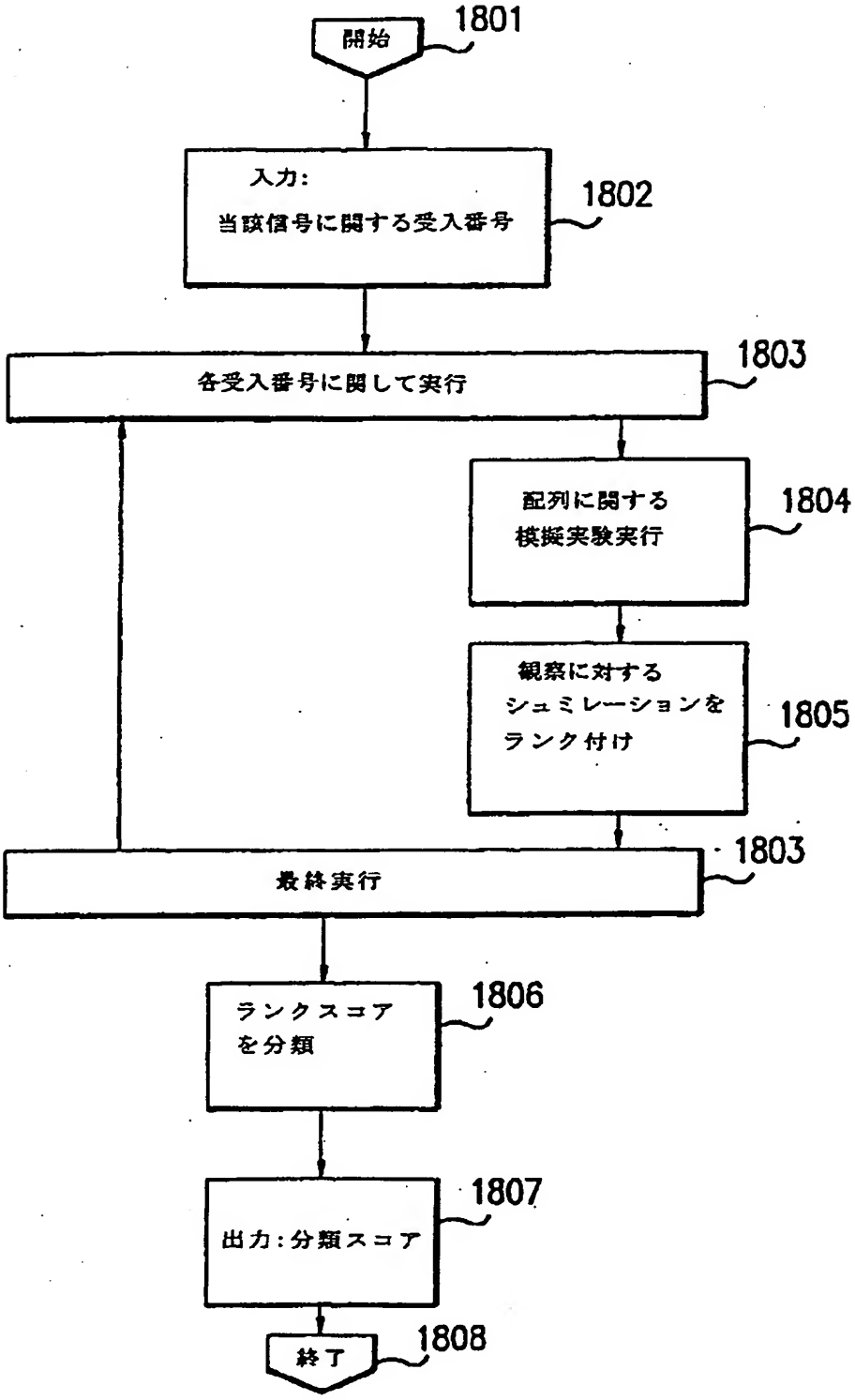


FIG.14

[Drawing 15]

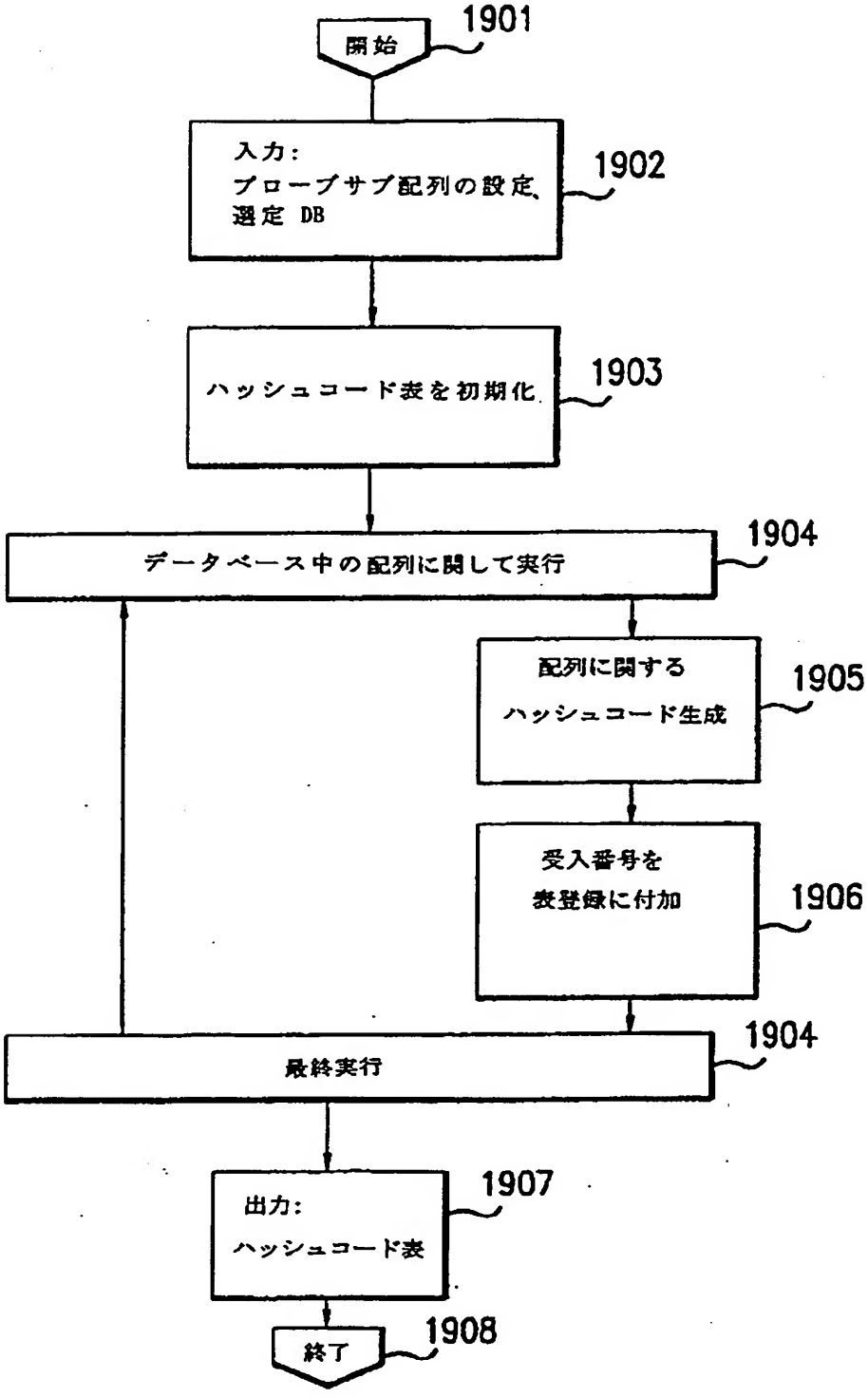


FIG.15

[Drawing 16]

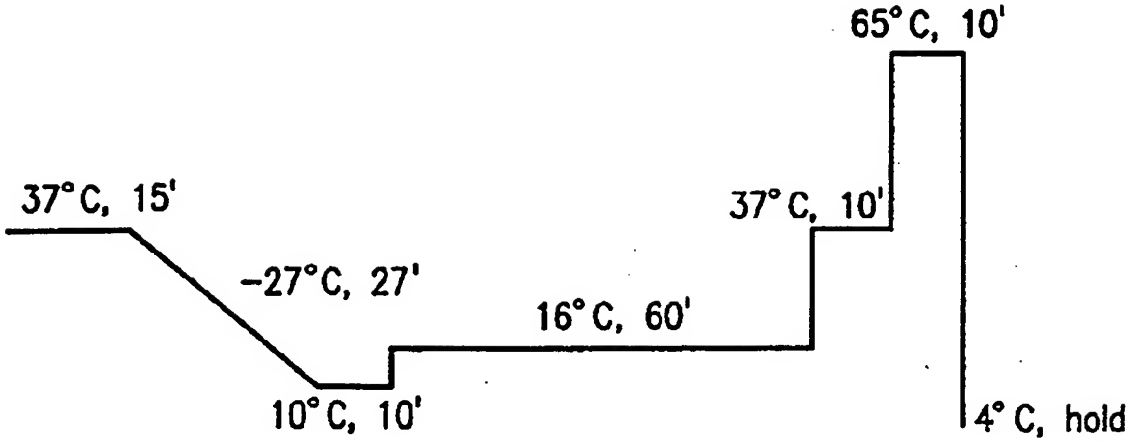


FIG.16A

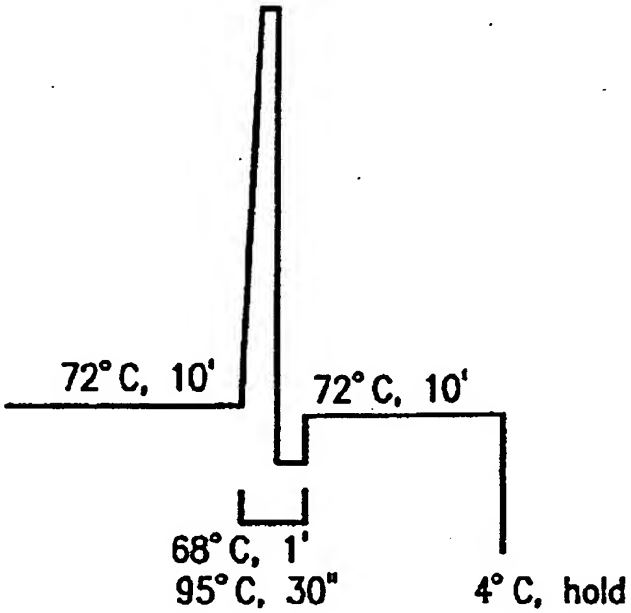


FIG.16B

[Drawing 16]

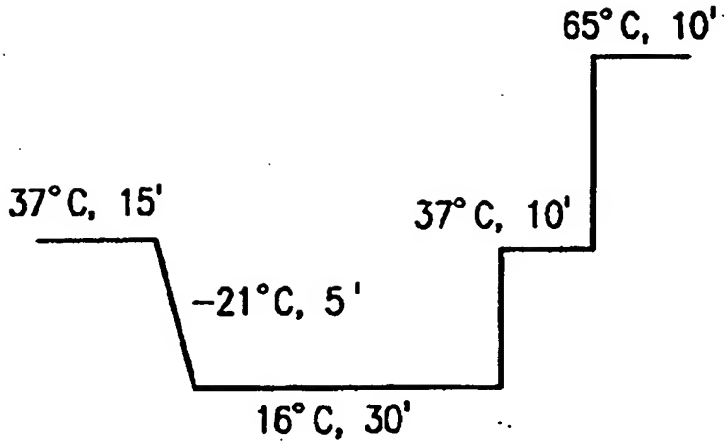


FIG.16C

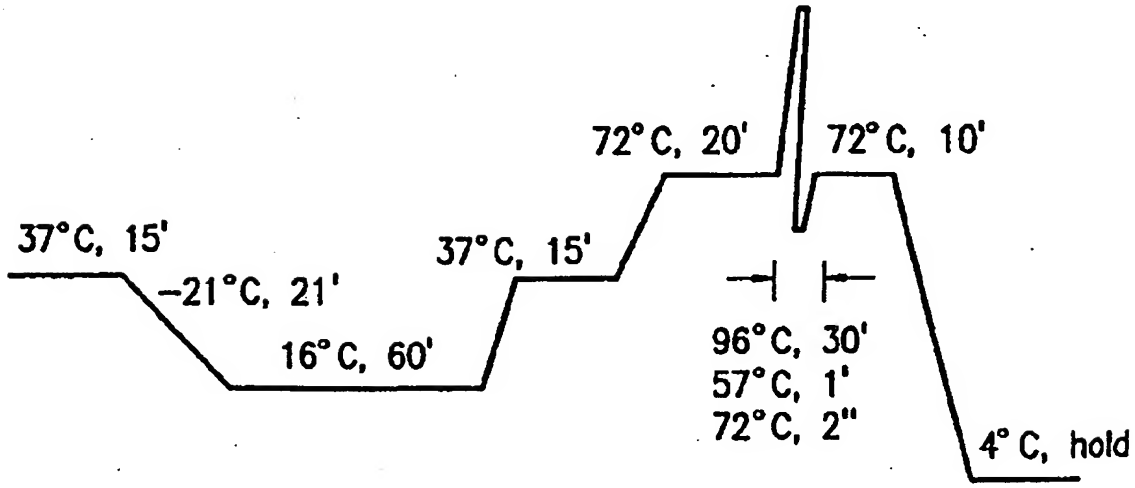


FIG.16D

[Drawing 17]

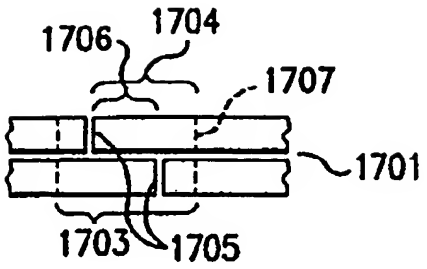


FIG.17A

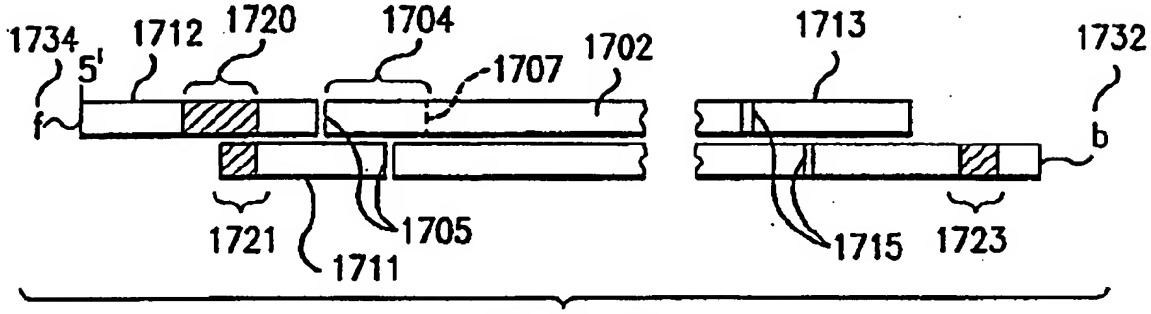


FIG.17B

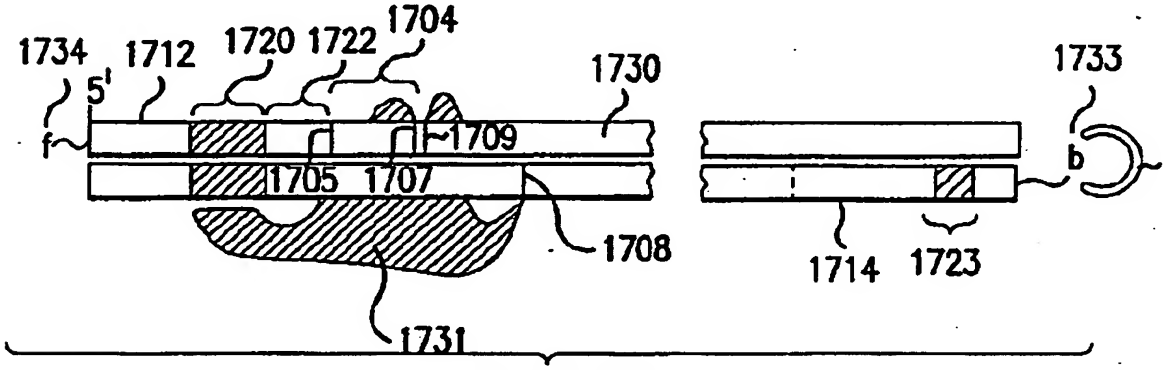


FIG.17C

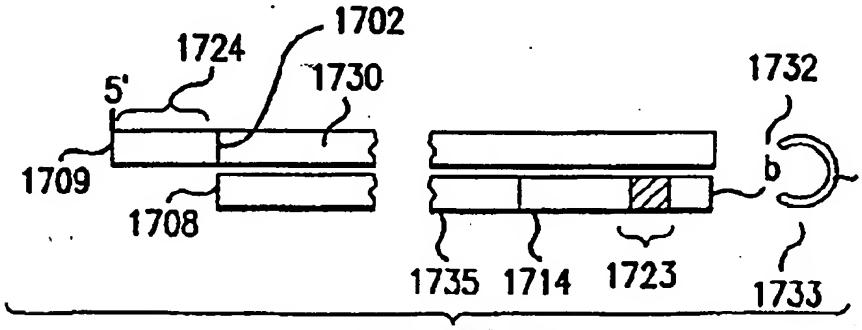


FIG.17D

[Drawing 17]

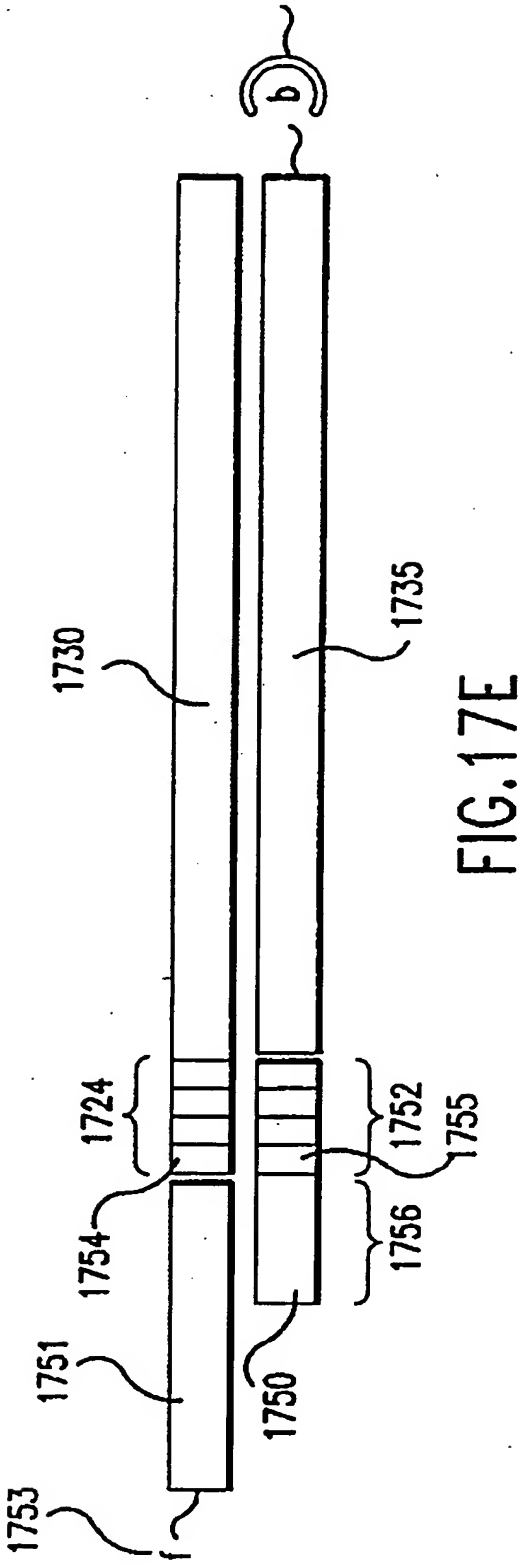


FIG.17E

[Translation done.]